The Annals of Statistics
2003, Vol. 31, No. 6, 1885–1904

ASYMPTOTIC RESULTS IN JACKKNIFING NONSMOOTH FUNCTIONS OF THE SAMPLE MEAN VECTOR

BY MARZIA MARCHESELLI

Università di Siena

The asymptotic behavior of jackknife estimators and jackknife variance estimators is investigated for nonsmooth functions of the sample mean vector. An application of jackknifing a suitable estimator of the intrinsic diversity profile is also presented.

1. Introduction. The jackknife method is widely applied in order to construct variance estimators and define bias-reduced estimators. Many results for the consistency of jackknife variance estimators are known for sufficiently smooth estimators $\hat{T}_n$ of T [Miller (1968), Reeds (1978), Parr and Schucany (1982) and Shao and Wu (1989)]. If estimators are not smooth, the jackknife may lead to an inconsistent variance estimator as shown in Miller (1974). Other results have been obtained for asymptotic behavior of jackknife bias reduction and jackknife estimators under suitable differentiability conditions [Miller (1964), Arvesen (1969), Thorburn (1977), Ghosh (1985) and Shao (1993)]. In this paper the consistency of jackknife variance estimators and the asymptotic behavior of jackknife estimators are studied for estimators $\hat{T}_n$ of the type $g(\overline{X}_n)$ when g is not necessarily a Gâteaux differentiable function but g satisfies some weaker regularity conditions which do not imply the asymptotic normality of $n^{1/2}(\hat{T}_n - T)$. Section 2 contains some preliminaries and notation while the asymptotic behavior of jackknife variance estimators and jackknife estimators is derived in Sections 3 and 4, respectively. Moreover, in Section 5, an asymptotically conservative confidence set for $g(\theta)$ is obtained, when g is not necessarily a differentiable function at θ, and, in this setting, the possible use of jackknife variance estimators is discussed. Section 6 focuses on these results as applied to an important family of nonsmooth estimators such as those obtained by estimating intrinsic diversity profiles.

2. Preliminaries and notation. Let $(X_n)_n$ be a sequence of i.i.d. random vectors, with values in a normed space $(V, \|\cdot\|)$ and with $E[\|X_1\|^2] < \infty$. Moreover, let g be a function defined on V, with values in a normed space $(W, |\!|\!|\cdot|\!|\!|)$, and let $\overline{X}_n$ be the sample mean $(X_1 + \cdots + X_n)/n$. The Quenouille–Tukey

Received February 2002; revised March 2003.
AMS 2000 subject classifications. Primary 62G05; secondary 62G20.
Key words and phrases. Delta method, jackknife, consistency, regular quasi-differentiability, intrinsic diversity profile.

(delete-1)jackknife estimator for $g(\overline{X}_n)$ is given by

$$J_n = \frac{1}{n}\sum_{k=1}^{n}\big[ng(\overline{X}_n) - (n-1)g(\overline{X}_{n,k})\big],$$

where $\overline{X}_{n,k}$ is the sample mean obtained by deleting the observation with the index k (obviously $k \le n$). When $W = \mathbb{R}$, the jackknife variance estimator can be defined for $g(\overline{X}_n)$ by

$$v_n = \frac{n-1}{n}\sum_{k=1}^{n}\left(g(\overline{X}_{n,k}) - \frac{1}{n}\sum_{j=1}^{n} g(\overline{X}_{n,j})\right)^2.$$

Now let θ be a vector of V and let $(A_i)_i$ be a finite partition of V. If A_i is a Borel set and $b(A_i - \theta) = A_i$ for any real number $b > 0$, the partition $(A_i)_i$ is said to be *regular* with respect to θ. Obviously, θ is an accumulation point for any set A_i of a regular partition.

DEFINITION 1. A function g is said to be *regularly quasi-differentiable* at θ if there exists a regular finite partition $(A_i)_i$ of V such that, for any i, g is A_i-differentiable at θ, namely there is a unique continuous linear operator L_i^θ such that

$$\lim_{v\to\theta, v\in A_i} \frac{|||g(v) - g(\theta) - L_i^\theta(v-\theta)|||}{\|v-\theta\|} = 0. \tag{1}$$

Any mapping τ (not necessarily linear on V) such that

$$\tau(v) = \sum_i L_i^\theta(v) I_{A_i}(v), \tag{2}$$

where L_i^θ denotes the A_i-differential of g at θ, is called *regular quasi-differential*.

The set of discontinuity points of τ is obviously a subset (in general, a proper subset) of $\bigcup_i \partial A_i$, for any regular partition $(A_i)_i$. However, it should be noticed that τ does not depend on the choice of the regular partition $(A_i)_i$ and if there exists a regular quasi-differential of g at θ, it is unique; that is, two regular quasi-differentials of g at θ coincide. A Fréchet differentiable function is obviously regularly quasi-differentiable while a quasi-differentiable function is Fréchet differentiable only if there exists a linear quasi-differential. Moreover, a regularly quasi-differentiable function g at θ is a Lipschitz function at θ; namely, there exists a constant C such that $|||g(v) - g(\theta)||| \le C\|v-\theta\|$. When $V = \mathbb{R}$, a function g is regularly quasi-differentiable at θ if and only if there exist finite $g'_+(\theta)$ and $g'_-(\theta)$. An interesting family of regularly quasi-differentiable functions is obtained when $g = h \circ \Gamma$, where $\Gamma : \mathbb{R}^k \to \mathbb{R}^k$ is the ordering function and h is a Fréchet differentiable function. When θ has at least two equal components,

$g = h \circ \Gamma$ is not necessarily Gâteaux differentiable at θ; nevertheless the regular quasi-differential of $h \circ \Gamma$ is $H \circ \widetilde{\Gamma}$, where H is the differential of h at $\Gamma(\theta)$ and $\widetilde{\Gamma}(x)$ is a suitable re-ordering of the components of x [for more details see Marcheselli (2000)]. By using the notion of regularly quasi-differentiable functions, in Marcheselli (2000) the following generalization of the delta method is given.

THEOREM. *Let Y_n, Y be random vectors in V and let $(a_n)_n$ represent a sequence of real numbers, with $\lim_n a_n = \infty$. Suppose that the random vector $a_n(Y_n - \theta)$ converges in distribution to Y. Moreover, let g be a regularly quasi-differentiable function at θ and suppose that $P(Y \in C) = 0$, where C is the set of discontinuity points of the regular quasi-differential τ. Then $a_n(g(Y_n) - g(\theta))$ converges in distribution to $\tau(Y)$.*

If V is a separable Hilbert space and R is a symmetric Gaussian random vector, with the same variance-covariance operator as X_1, from the previous theorem it follows that

$$\sqrt{n}\big(g(\overline{X}_n) - g(\theta)\big) \tag{3}$$

converges in distribution to $\tau(R)$, when g is a regularly quasi-differentiable function at $\theta = E[X_1]$ and $P(R \in C) = 0$, where C is the set of discontinuity points of τ. Note that $\tau(R)$ is not necessarily a Gaussian random vector. Moreover, $E[\tau(R)] = 0$ when g is differentiable at θ but, in general, $E[\tau(R)] \neq 0$ and, since g is a Lipschitz function at θ, the bias $E[g(\overline{X}_n)] - g(\theta)$ is equivalent to $E[\tau(R)]/\sqrt{n}$ and $\mathrm{Var}[g(\overline{X}_n)]$ is equivalent to $\mathrm{Var}[\tau(R)]/n$. Therefore, the classic techniques cannot be used in order to study the asymptotic behavior of jackknife estimators for $g(\overline{X}_n)$, when g is only a regularly quasi-differentiable function at θ. To this end, the following definition is also required.

DEFINITION 2. A function g is said to be *continuously regularly quasi-differentiable* at θ if g is regularly quasi-differentiable at θ and there exists a regular partition $(A_i)_i$ such that, for any index i:

(a) A_i is a convex set;

(b) there exists a convex neighborhood $\mathcal{U}$ of θ such that, for any element $x \in A_i \cap \mathcal{U}$, g is A_i-differentiable at x and

$$\lim_{x \to \theta, x \in A_i} \|L_i^x - L_i^\theta\|_{\mathcal{L}(V,W)} = 0,$$

where L_i^x denotes the A_i-differential of g at x and $\mathcal{L}(V, W)$ is the space of continuous linear operators, defined on V and with values in W.

It is apparent that a continuously Fréchet differentiable function at θ is also continuously regularly quasi-differentiable at θ. Functions g of the type $h \circ \Gamma$, with h a continuously differentiable function, are obviously continuously regularly quasi-differentiable but, in general, are not continuously differentiable.

3. Consistency of jackknife variance estimators. Let V be a separable Hilbert space, R be a symmetric Gaussian random vector with the same variance-covariance operator as X_1 and $\theta = E[X_1]$. The jackknife variance estimator v_n is often adopted since it has good asymptotic performance and its computation only requires n evaluations of g at $\overline{X}_{n,j}$, $j = 1, \dots, n$, which is routine and easy to program. It is well known that v_n is strongly consistent when $g : \mathbb{R}^k \to \mathbb{R}$ is continuously differentiable at θ [see Shao and Tu (1995)], namely, $n v_n$ converges almost surely to the variance of $\tau(R) = \langle \nabla g(\theta), R \rangle$. In this section, some results on the asymptotic behavior of v_n are derived when g satisfies some weaker smoothness conditions than continuous differentiability at θ.

LEMMA 1. *Let $(A_i)_i$ be a regular partition of V with respect to θ. If $P(R \in \bigcup_i \partial A_i) = 0$,*

$$U_n = \sum_i \sum_{k=1}^{n} \left| I_{A_i}(\overline{X}_{n,k}) - I_{A_i}(\overline{X}_n) \right|$$

converges in probability to 0.

PROOF. Obviously U_n converges in probability to 0 when X_1 is almost surely a constant. It is not restrictive to suppose that $X_1(P)$ is not degenerate. Thus, it is sufficient to prove that, for any i,

$$U_n^i = \sum_{k=1}^{n} \left| I_{A_i}(\overline{X}_{n,k}) - I_{A_i}(\overline{X}_n) \right|$$

converges in probability to 0. Since $b(A_i - \theta) = A_i$ for any real number $b > 0$, it follows that

$$\begin{aligned} U_n^i &= I_{A_i}(\overline{X}_n) \sum_{k=1}^{n} I_{A_i^c}(\overline{X}_{n,k}) + I_{A_i^c}(\overline{X}_n) \sum_{k=1}^{n} I_{A_i}(\overline{X}_{n,k}) \\ &= I_{A_i}(Y_n) \sum_{k=1}^{n} I_{A_i^c}(Y_{n,k}) + I_{A_i^c}(Y_n) \sum_{k=1}^{n} I_{A_i}(Y_{n,k}), \end{aligned}$$

where Y_n and $Y_{n,k}$ denote the random variables $\sqrt{n}(\overline{X}_n - \theta)$ and $\sqrt{n}(\overline{X}_{n,k} - \theta)$, respectively. For a fixed real number $\delta > 0$, since $\|Y_n - Y_{n,k}\| \geq 2\delta$ implies that $\|Y_n\|/(n-1) + 2\|X_k - \theta\|/\sqrt{n} \geq 2\delta$, the following relation holds:

$$\{U_n^i \neq 0\} \subseteq \{\|Y_n\| \geq (n-1)\delta\} \cup \left\{ \sup_{1 \leq k \leq n} \frac{\|X_k - \theta\|}{\sqrt{n}} \geq \delta/2 \right\} \cup \bigcup_j \{Y_n \in A_j^{2\delta}\},$$

where $A_j^{2\delta}$ is the open set $\partial A_j + B(0, 2\delta)$, with $B(0, 2\delta) = \{x : \|x\| < 2\delta\}$. Since $\|X_1 - \theta\|^2$ is integrable, the random variable $\sup_{1 \leq k \leq n} \|X_k - \theta\|/\sqrt{n}$ converges almost surely to 0. Moreover, owing to the central limit theorem, Y_n converges in distribution to R [see Varadhan (1962)]. Then it follows that

$$\limsup_n P(U_n^i \neq 0) \leq \sum_j P(R \in \bar{A}_j^{2\delta}),$$

where $\bar{A}_j^{2\delta}$ is the closure of $A_j^{2\delta}$. Finally, the relation

$$\lim_{\delta \to 0^+} P(R \in \bar{A}_j^{2\delta}) = P(R \in \partial A_j) = 0$$

completes the proof. □

From now on, if $(Z_n)_n$ and $(W_n)_n$ are two sequences of real random variables, $Z_n \sim W_n$ means that Z_n converges in probability if and only if W_n converges in probability and they have the same limit. The following convergence result for variance jackknife estimators can now be proved.

THEOREM 1. *Let $g : V \to \mathbb{R}$ be a continuously regularly quasi-differentiable function at θ. Let $(A_i)_i$ be a regular partition with respect to θ which satisfies* (a) *and* (b) *of Definition* 2. *Suppose that $P(R \in \partial A_i) = 0$ for any index i. Then nv_n converges in distribution to $\sum_i I_{A_i}(R)\,\mathrm{Var}[L_i^\theta(X_1)]$. Moreover, nv_n converges in probability if and only if there exists a constant c such that $\mathrm{Var}[L_i^\theta(X_1)] = c$ for any i, and, in this case, nv_n converges in probability to c.*

PROOF. According to the definition of jackknife variance estimators it follows that

$$\begin{aligned} nv_n &= (n-1)\sum_{k=1}^n \left(g(\overline{X}_{n,k}) - \frac{1}{n}\sum_{j=1}^n g(\overline{X}_{n,j}) \right)^2 \\ &= (n-1)\sum_{k=1}^n \left(g(\overline{X}_{n,k}) - g(\overline{X}_n) - \frac{1}{n}\sum_{j=1}^n [g(\overline{X}_{n,j}) - g(\overline{X}_n)] \right)^2 \\ &\sim n\sum_{k=1}^n \big(g(\overline{X}_{n,k}) - g(\overline{X}_n)\big)^2 - \left[\sum_{k=1}^n \big(g(\overline{X}_n) - g(\overline{X}_{n,k})\big) \right]^2. \end{aligned}$$

Now, let $\mathcal{U}$ be a convex neighborhood of θ which satisfies condition (b) of Definition 2. Having fixed $\delta > 0$, it is not restrictive to suppose $\mathcal{U}_\delta = B(\theta, \delta) \subset \mathcal{U}$. Since

$$I_{\mathcal{U}_\delta^c}(\overline{X}_n) + \sup_k I_{\mathcal{U}_\delta^c}(\overline{X}_{n,k})$$

converges almost surely to 0, in order to study the asymptotic behavior of nv_n, note that $g(\overline{X}_{n,k}) - g(\overline{X}_n)$ is equivalent to $\sum_i (g(\overline{X}_{n,k}) - g(\overline{X}_n))[I_{A_i \cap \mathcal{U}_\delta}(\overline{X}_{n,k}) \times I_{A_i \cap \mathcal{U}_\delta}(\overline{X}_n) + |I_{A_i}(\overline{X}_{n,k}) - I_{A_i}(\overline{X}_n)|/2]$ and

$$n \sum_{k=1}^{n} \big(g(\overline{X}_{n,k}) - g(\overline{X}_n)\big)^2 \sim n \sum_i \sum_{k=1}^{n} \big(g(\overline{X}_{n,k}) - g(\overline{X}_n)\big)^2 I_{A_i \cap \mathcal{U}_\delta}(\overline{X}_{n,k}) I_{A_i \cap \mathcal{U}_\delta}(\overline{X}_n) + C_n,$$

where $C_n = n \sum_i \sum_{k=1}^n (g(\overline{X}_{n,k}) - g(\overline{X}_n))^2 |I_{A_i}(\overline{X}_{n,k}) - I_{A_i}(\overline{X}_n)|/2$. Then, from Lemma 1 and from the hypothesis of g, it follows that C_n converges in probability to 0 and

$$n \sum_{k=1}^{n} \big(g(\overline{X}_{n,k}) - g(\overline{X}_n)\big)^2 \sim I_{\mathcal{U}_\delta}(\overline{X}_n) n \sum_i \sum_{k=1}^{n} [L_i^{\xi_{k,i}}(\overline{X}_{n,k} - \overline{X}_n)]^2 I_{A_i}(\overline{X}_n),$$

where $\xi_{k,i}$ is a suitable random variable with values in $A_i \cap \mathcal{U}_\delta$. Since $\overline{X}_{n,k} - \overline{X}_n = [(\overline{X}_n - \theta) + (\theta - X_k)]/(n-1)$, it may be observed that

$$\begin{aligned} n \sum_i \sum_{k=1}^{n} [(L_i^{\xi_{k,i}} - L_i^\theta)(\overline{X}_{n,k} - \overline{X}_n)]^2 &\le n \sum_i \Big(\sup_{x \in A_i \cap \mathcal{U}_\delta} \|L_i^x - L_i^\theta\|^2_{\mathcal{L}(V,\mathbb{R})} \Big) \sum_{k=1}^{n} \|\overline{X}_{n,k} - \overline{X}_n\|^2 \\ &\sim E[\|X_1 - \theta\|^2] \sum_i \Big(\sup_{x \in A_i \cap \mathcal{U}_\delta} \|L_i^x - L_i^\theta\|^2_{\mathcal{L}(V,\mathbb{R})} \Big), \end{aligned}$$

that is, $n \sum_i \sum_{k=1}^n [(L_i^{\xi_{k,i}} - L_i^\theta)(\overline{X}_{n,k} - \overline{X}_n)]^2$ converges almost surely to 0 owing to the continuous regular quasi-differentiability of g at θ. From the previous relation, it follows that

$$\begin{aligned} n \sum_{k=1}^{n} \big(g(\overline{X}_{n,k}) - g(\overline{X}_n)\big)^2 &\sim I_{\mathcal{U}_\delta}(\overline{X}_n) n \sum_i \sum_{k=1}^{n} [L_i^\theta(\overline{X}_{n,k} - \overline{X}_n)]^2 I_{A_i}(\overline{X}_n) \\ &\sim \sum_i I_{A_i}(\overline{X}_n) n \sum_{k=1}^{n} [L_i^\theta(\overline{X}_{n,k} - \overline{X}_n)]^2 \\ &= \sum_i I_{A_i}\big(\sqrt{n}(\overline{X}_n - \theta)\big) n \sum_{k=1}^{n} [L_i^\theta(\overline{X}_{n,k} - \overline{X}_n)]^2. \end{aligned}$$

In a similar way, it turns out that

$$\left[\sum_{k=1}^{n}\big(g(\overline{X}_n)-g(\overline{X}_{n,k})\big)\right]^2 \sim \left(\sum_i\sum_{k=1}^{n}[L_i^\theta(\overline{X}_{n,k}-\overline{X}_n)]I_{A_i}(\overline{X}_n)\right)^2$$

$$=\left(\sum_i I_{A_i}(\overline{X}_n)L_i^\theta\left[\sum_{k=1}^{n}(\overline{X}_{n,k}-\overline{X}_n)\right]\right)^2=0.$$

Therefore nv_n is asymptotically equivalent in probability to

$$T_n=\sum_i I_{A_i}\big(\sqrt{n}(\overline{X}_n-\theta)\big)n\sum_{k=1}^{n}[L_i^\theta(\overline{X}_{n,k}-\overline{X}_n)]^2.$$

Owing to the strong law of large numbers, it is easy to recognize that

$$T_n\sim\sum_i I_{A_i}\big(\sqrt{n}(\overline{X}_n-\theta)\big)\sum_{k=1}^{n}[L_i^\theta(X_k-\theta)]^2/n$$

$$\sim\sum_i I_{A_i}\big(\sqrt{n}(\overline{X}_n-\theta)\big)E[L_i^\theta(X_1-\theta)^2].$$

Observe that $E[L_i^\theta(X_1-\theta)^2]=\mathrm{Var}[L_i^\theta(X_1)]$. Thanks to the central limit theorem and the negligibility of ∂A_i, T_n and nv_n converge in distribution to $L=\sum_i I_{A_i}(R)\,\mathrm{Var}[L_i^\theta(X_1)]$. Moreover, nv_n converges in probability if and only if L is a constant; namely, if there exists a real number c such that $\mathrm{Var}[L_i^\theta(X_1)]=c$, for any i. Thus the proof is complete. □

REMARK 1. When g is a continuously regularly quasi-differentiable function at θ, it may occur that nv_n does not converge either to the second moment or to the variance of the limit random variable $\tau(R)$. For example, if

$$g(x_1,x_2)=x_1\wedge x_2,\qquad X_1=(X_1',X_1''),\qquad \theta=E[X_1]=(0,0)$$

and X_1' and $X_1''-X_1'$ are not degenerate independent random variables, by considering the regular partition $(A_i)_{i=1,2}$ of $\mathbb{R}^2$, with

$$A_1=\{x_1\le x_2\},\qquad A_2=\{x_1>x_2\},\tag{4}$$

it turns out that $\tau(R)=g(R_1,R_2)$. Since the hypotheses of Theorem 1 are satisfied, nv_n converges in distribution to $L=\mathrm{Var}[X_1']+I_{\{R_1>R_2\}}\mathrm{Var}[X_1''-X_1']$. In this case, nv_n does not converge in probability and its distribution-limit L does not coincide with the variance (nor with the second moment) of $\tau(R)$! It may only be stated that $E[L]=E[\tau^2(R)]$. Moreover, since g satisfies the Lipschitz condition at θ, it should be noticed that the bias $E[g(\overline{X}_n)]-g(\theta)$ is equivalent to $E[\tau(R)]/\sqrt{n}$, which obviously is not $o(1/\sqrt{n})$, and $E[(g(\overline{X}_n)-g(\theta))^2]$ is equivalent to $E[\tau^2(R)]/n$.

REMARK 2. In Lemma 1, U_n converges in probability to 0 but it does not necessarily converge in mean [to this purpose it suffices to consider U_n when X_1 is a bivariate Gaussian random vector and $(A_i)_i$ is given by (4)]. Moreover, from the proof of Theorem 1, it is easy to note that if U_n converges almost surely to 0, nv_n converges in probability if and only if nv_n almost surely converges.

As shown in Remark 1, nv_n may be an inconsistent estimator of the variance of $\tau(R)$ when g is continuously regularly quasi-differentiable at θ. Some examples in which nv_n converges in probability but its limit is not related to the moments of $\tau(R)$ could be also introduced. Note that the bootstrap variance estimator for $g(\overline{X}_n)$ may be an inconsistent estimator of $\mathrm{Var}[\tau(R)]$; in fact, it has the same asymptotic behavior as nv_n since, on every event $\{\overline{X}_n \in A_i\}$, the assumptions of regularity for g are equivalent to their classical counterparts. Nevertheless, under suitable conditions, the jackknife variance nv_n converges in probability to the second moment of $\tau(R)$. Actually, the following result can be proved.

COROLLARY 1. *Under the hypotheses of Theorem* 1, *let H be a closed subspace of V. Denote by π_H the orthogonal projection onto H and let a be a real number. For any index i, suppose that $L_i^\theta(\pi_H(R))$ is independent of $I_{A_i}(R), L_i^\theta(\pi_K(R))$, where K denotes the orthogonal subspace of H, and $E[L_i^\theta(\pi_H(R))^2] = a$. If $|L_i^\theta|$ are coincident on K, then nv_n converges in probability to $E[\tau(R)^2]$.*

PROOF. First, note that, for any index i and for $J = \{R \in A_i\}$ or $J = \Omega$,

$$\begin{aligned} E[I_J L_i^\theta(R)^2] &= E\big[I_J L_i^\theta\big(\pi_H(R)\big)^2\big] + E\big[I_J L_i^\theta\big(\pi_K(R)\big)^2\big] \\ &= E\big[I_J L_i^\theta\big(\pi_H(R)\big)^2\big] + E\big[I_J L_1^\theta\big(\pi_K(R)\big)^2\big]. \end{aligned}$$

Since $\mathrm{Var}[L_i^\theta(X_1)] = \mathrm{Var}[L_i^\theta(R)]$, from the previous relation, if $J = \Omega$, it follows that

$$\mathrm{Var}[L_i^\theta(X_1)] = a + E\big[L_1^\theta\big(\pi_K(R)\big)^2\big]$$

and then, thanks to Theorem 1, nv_n converges in probability. Moreover,

$$\begin{aligned} E[\tau(R)^2] &= \sum_i E[I_{A_i}(R) L_i^\theta(R)^2] \\ &= \sum_i P(A_i \in R)a + E\big[I_{A_i}(R) L_1^\theta\big(\pi_K(R)\big)^2\big] = a + E\big[L_1^\theta\big(\pi_K(R)\big)^2\big], \end{aligned}$$

that is, nv_n converges in probability to the second moment of $\tau(R)$. The corollary is thus proved. □

EXAMPLE. Let

$$g(x_1, x_2) = \frac{x_1 \wedge x_2}{x_1 + x_2} I_{\{x_1+x_2 \neq 0\}}, \qquad X_1 = (X_1', X_1''),$$
$$\theta = E[X_1] = (\theta_1, \theta_1) \neq (0, 0),$$

with $P(X_1' \neq X_1'') > 0$. By considering the regular partition of $\mathbb{R}^2$ given in (4), g is a continuously quasi-differentiable function, but not Gâteaux differentiable at θ and, owing to the generalized delta method, $\tau(R) = -|R_2 - R_1|/(4\theta_1)$ is the limit in distribution of $\sqrt{n}(g(\overline{X}_n) - g(\theta))$. Since $|L_1^\theta| = |L_2^\theta|$, thanks to Corollary 1, nv_n converges in probability to $E[\tau(R)^2]$. In particular, whatever the variance-covariance matrix of X_1 may be, $(g(\overline{X}_n) - g(\theta))/\sqrt{v_n}$ converges in distribution to $-|Z| \operatorname{sgn}(\theta_1)$, where Z is a standard Gaussian random variable, and $E[(g(\overline{X}_n) - g(\theta))^2]/v_n$ converges in probability to 1.

4. Asymptotic behavior of jackknife estimators. The asymptotic behavior of J_n is known when g is twice continuously differentiable at θ. In this section, the asymptotic distribution of J_n is obtained when g satisfies some weaker regularity conditions. To this purpose it is necessary to introduce a definition which generalizes the concept of continuous differentials of the second order in a natural way.

DEFINITION 3. A function g is said to be *twice continuously regularly quasi-differentiable* at θ if there exist a regular partition $(A_i)_i$ and a convex neighborhood $\mathcal{U}$ of θ such that g is continuously regularly quasi-differentiable at θ and, for any index i, the mapping $L_i : x \mapsto L_i^x$ is A_i-differentiable at every point of $A_i \cap \mathcal{U}$ and

$$(*) \qquad \lim_{x \to \theta, x \in A_i} \|\beta_i^x - \beta_i^\theta\|_{\mathcal{L}(V, \mathcal{L}(V,W))} = 0,$$

where β_i^x denotes the A_i-differential of L_i at x and $\mathcal{L}(V, \mathcal{L}(V, W))$ is the space of continuous linear operators, defined on V and with values in $\mathcal{L}(V, W)$.

Roughly speaking, a function g is twice continuously regularly quasi-differentiable if there exists a finite partition $(A_i)_i$ such that g is twice continuously differentiable on every A_i. By using the same notation as in the previous sections, the following result can now be proved.

THEOREM 2. *Let g be a twice continuously regularly quasi-differentiable function at θ and let $(A_i)_i$ be a regular partition with respect to θ which verifies* $(*)$ *and* (a), (b) *of Definition* 2. *Suppose that $P(R \in \partial A_i) = 0$, for any index i. Then*

$$M_n = \sqrt{n}\big(J_n - g(\overline{X}_n)\big)$$

converges in probability to 0. *Moreover*, $\sqrt{n}[J_n - g(\theta)]$ *converges in distribution to* $\tau(R)$.

PROOF. By means of simple algebraic relations, it turns out that

$$\begin{aligned}\sqrt{n}\big(J_n - g(\overline{X}_n)\big) &= \frac{1}{\sqrt{n}}\sum_{k=1}^{n}\big[ng(\overline{X}_n) - (n-1)g(\overline{X}_{n,k}) - g(\overline{X}_n)\big]\\ &= \frac{n-1}{\sqrt{n}}\sum_{k=1}^{n}[g(\overline{X}_n) - g(\overline{X}_{n,k})]\\ &\sim \sqrt{n}\sum_{k=1}^{n}[g(\overline{X}_n) - g(\overline{X}_{n,k})].\end{aligned}$$

Since $I_{\mathcal{U}^c}(\overline{X}_n) + \sup_k I_{\mathcal{U}^c}(\overline{X}_{n,k})$ converges almost surely to 0, from the definition of a twice continuously regularly quasi-differentiable function at θ and from Lemma 1, it follows that

$$\begin{aligned}\Delta_n &= \sum_{k=1}^{n}[g(\overline{X}_n) - g(\overline{X}_{n,k})]\\ &\sim \sum_i\sum_{k=1}^{n} L_i^{\overline{X}_n}(\overline{X}_n - \overline{X}_{n,k}) I_{A_i\cap\mathcal{U}}(\overline{X}_n)\\ &\quad - \tfrac{1}{2}\sum_i\sum_{k=1}^{n}\langle\beta_i^{\eta_{i,k}}(\overline{X}_n - \overline{X}_{n,k}), (\overline{X}_n - \overline{X}_{n,k})\rangle I_{A_i\cap\mathcal{U}}(\overline{X}_n),\end{aligned}$$

where $\eta_{i,k}$ is a suitable random variable with values in $A_i \cap \mathcal{U}$. Thus $\sqrt{n}(J_n - g(\overline{X}_n))$ may be equivalently decomposed into the sum of two components γ_n,ε_n, where

$$\gamma_n = \sum_i \sqrt{n}\sum_{k=1}^{n} L_i^{\overline{X}_n}(\overline{X}_n - \overline{X}_{n,k}) I_{A_i\cap\mathcal{U}}(\overline{X}_n),$$

$$\varepsilon_n = -\tfrac{1}{2}\sum_i \sqrt{n}\sum_{k=1}^{n}\langle\beta_i^{\eta_{i,k}}(\overline{X}_n - \overline{X}_{n,k}), (\overline{X}_n - \overline{X}_{n,k})\rangle I_{A_i\cap\mathcal{U}}(\overline{X}_n).$$

Now, γ_n, ε_n remain to be studied. It may first be observed that

$$\begin{aligned}\gamma_n &= \sum_i I_{A_i\cap\mathcal{U}}(\overline{X}_n)\sum_{k=1}^{n}\sqrt{n}L_i^{\overline{X}_n}(\overline{X}_n - \overline{X}_{n,k})\\ &= \sum_i I_{A_i\cap\mathcal{U}}(\overline{X}_n)\sqrt{n}L_i^{\overline{X}_n}\left(\sum_{k=1}^{n}\overline{X}_n - \overline{X}_{n,k}\right) = 0\end{aligned}$$

since $\sum_{k=1}^{n}(\overline{X}_n - \overline{X}_{n,k}) = 0$. In order to prove that ε_n converges in probability to 0, it suffices to show that

$$\delta_n = \sqrt{n}\sum_{k=1}^{n}\|\overline{X}_n - \overline{X}_{n,k}\|^2$$

converges in probability to 0. Since

$$\delta_n \le 2\frac{\sqrt{n}}{(n-1)}\left(2\|\overline{X}_n\|^2 + \sum_{k=1}^{n}\|X_k\|^2/(n-1)\right),$$

δ_n converges in probability (and almost surely) to 0. The proof is therefore complete. □

REMARK 3. In general, the ratio $r_n = (E[J_n] - g(\theta))/(E[g(\overline{X}_n)] - g(\theta))$ is not negligible when g is only twice continuously regularly quasi-differentiable at θ. Sometimes r_n is equivalent to a constant $c \neq 0$. For example, with reference to the case of Remark 1, if X_1 is a bivariate Gaussian random vector, it is not difficult to prove that $E[M_n] = E[\sqrt{n}(J_n - g(\overline{X}_n))]$ converges to $E[(X_1'' - X_1')^+]/2$ [which coincides with $-E[\tau(R)]/2$] and then r_n converges to $c = 1/2$. As a matter of fact, Lemma (3.2) in Marcheselli (1999) shows that

$$M_n = \frac{1}{\sqrt{n}}\sum_{k=1}^{n}|W_{n,k}|\big(I_{\{W_n<0<W_{n,k}\}} + I_{\{W_{n,k}<0\le W_n\}}\big),$$

where $W_n = (X_1'' - X_1') + \cdots + (X_n'' - X_n')$ and $W_{n,k} = W_n - (X_k'' - X_k')$. In particular, both M_n and $nb_{\mathrm{Jack}} = -\sqrt{n}M_n$ converge in probability to 0 but not in L^1. Then, unlike what happens if g is continuously twice differentiable at θ, the bias of the jackknife estimator J_n is not considerably reduced with respect to the bias of $g(\overline{X}_n)$. Interesting cases where J_n does not reduce the bias may be obtained by considering twice continuously regularly quasi-differentiable functions g of the type $h \circ \Gamma$ when g is not differentiable at θ.

5. Asymptotically conservative confidence sets.

By using the same notation as in the previous sections, let $g : \mathbb{R}^k \to \mathbb{R}^m$ be a regularly quasi-differentiable function at $\theta = E[X_1]$. Moreover, let $(A_i)_{i\in I}$ be a regular finite partition of $\mathbb{R}^k$ such that the regular quasi-differential τ is given by

$$\tau = \sum_{i\in I} L_i^\theta I_{A_i}$$

and, for any $j = 1, \dots, m$, let $I_j \subset I$ be an index set for the distinct functions $\pi_j \circ L_i^\theta$, $i \in I$, where π_j denotes the orthogonal projection onto the jth coordinate. Suppose that $\tau \not\equiv 0$ and that $P(R \in C) = 0$, where R is a symmetric Gaussian random vector on $\mathbb{R}^k$, with the same variance-covariance matrix Σ as X_1, and C is the set of discontinuity points of τ. Owing to the theorem in Section 2, $\tau(R)$ is the limit in distribution of $\sqrt{n}(g(\overline{X}_n) - g(\theta))$ and, as has already been observed, $\tau(R)$ is not a Gaussian vector if g is not differentiable at θ. In this case, neither the knowledge of the variance-covariance of $\tau(R)$, nor the knowledge of Σ, characterizes the distribution of $\tau(R)$ and, for $j = 1, \dots, m$, the jackknife variance estimators for $\pi_j \circ g(\overline{X}_n)$ are not consistent estimators either of the variance or of the second moment of $\pi_j \circ \tau(R)$ [notwithstanding that $\lim_n nE[(\pi_j \circ g(\overline{X}_n) - \pi_j \circ g(\theta))^2] = E[(\pi_j \circ \tau(R))^2]$. For this reason, the nondifferentiability of g at θ is a drawback to construction of asymptotic confidence sets for $g(\theta)$ by using the jackknife method on the components $\pi_j \circ g$.

However, thanks to a suitable application of the Šidák inequality (1967), asymptotically conservative confidence sets for $g(\theta)$ may be obtained by generalizing the Richmond method (1982). More precisely:

PROPOSITION 1. *Let Z be a standard Gaussian random variable and let b_i be the $m \times k$ matrix such that $L_i^\theta(v) = b_i v$ for any $v \in \mathbb{R}^k$. Then, for any $(z_1, \dots, z_m) \in \mathbb{R}_+^m$, it turns out that*

$$\lim_n P\left(\bigcap_{j=1}^m \sqrt{n}|g_j(\overline{X}_n) - g_j(\theta)| \le z_j\right) \ge \prod_{j=1}^m \prod_{i \in I_j} P(\sigma_{i,j}|Z| \le z_j),$$

where $g_j = \pi_j \circ g$ and $\sigma_{i,j} = \sqrt{(b_i \Sigma b_i^t)_{j,j}}$.

PROOF. From the definition of τ it follows that

$$\begin{aligned}\bigcap_{j=1}^m \{|\tau_j(R)| \le z_j\} &= \bigcap_{j=1}^m \bigcup_{i \in I} A_i \cap \{|\pi_j \circ L_i^\theta(R)| \le z_j\} \\ &\supseteq \bigcap_{j=1}^m \bigcap_{i \in I} \{|\pi_j \circ L_i^\theta(R)| \le z_j\} \\ &= \bigcap_{j=1}^m \bigcap_{i \in I_j} \{|\pi_j \circ L_i^\theta(R)| \le z_j\}.\end{aligned}$$

Moreover, since the random variables $\pi_j \circ L_i^\theta(R)$ are the components of a symmetric Gaussian random vector, owing to the Šidák inequality [Šidák (1967), Corollary 4, page 628], we may observe that

$$\begin{aligned}P\left(\bigcap_{j=1}^m \{|\tau_j(R)| \le z_j\}\right) &\ge P\left(\bigcap_{j=1}^m \bigcap_{i \in I_j} \{|\pi_j \circ L_i^\theta(R)| \le z_j\}\right) \\ &\ge \prod_{j=1}^m \prod_{i \in I_j} P\big(|\pi_j \circ L_i^\theta(R)| \le z_j\big).\end{aligned}$$

Finally, since $\mathrm{Var}[\pi_j \circ L_i^\theta(R)] = \sigma_{i,j}^2$, from the convergence in distribution of $\sqrt{n}(g(\overline{X}_n) - g(\theta))$ to $\tau(R)$, the proposition is straightforwardly proven. $\square$

Now, let us consider a particular case in which the regularly quasi-differentiable function g is of type $h \circ \Gamma$, where h is a Fréchet differentiable function with values in $\mathbb{R}^m$ and $\Gamma : \mathbb{R}^k \to \mathbb{R}^k$ is the decreasing ordering function. If $\{\vartheta_1, \dots, \vartheta_s\}$ is the set of the distinct components of θ, with $\vartheta_1 > \cdots > \vartheta_s$, let $F_1, \dots, F_s$ be the disjoint subsets of $\{1, \dots, k\}$ such that $F_h = \{j : \theta_j = \vartheta_h\}$, where $h = 1, \dots, s$. Moreover, let d_h be the cardinality of F_h and let $D_h =]d_1 + \cdots + d_{h-1}, d_1 + \cdots + d_h] \cap \{1, \dots, k\}$. Let I_j be a set of one-to-one functions ϕ on $\{1, \dots, k\}$, with $\phi(D_h) = F_h$, such that I_j is an index set for the distinct functions of the type $x \mapsto \langle \nabla h_j(\Gamma(\theta)), x_\phi \rangle$, where $x_\phi = (x_{\phi(1)}, \dots, x_{\phi(k)})$. Then the following result holds.

COROLLARY 2. *Let $g = (h_1 \circ \Gamma, \dots, h_m \circ \Gamma)$. If Z is a standard Gaussian random variable, for any $(z_1, \dots, z_m) \in \mathbb{R}^m_+$, it turns out that*

$$\lim_n P\left(\bigcap_{j=1}^m \sqrt{n} |h_j \circ \Gamma(\overline{X}_n) - h_j \circ \Gamma(\theta)| \le z_j \right) \ge \prod_{j=1}^m \prod_{\phi \in I_j} P(\sigma_{\phi,j}|Z| \le z_j), \tag{5}$$

where $\sigma_{\phi,j} = \sqrt{E[\langle \nabla h_j(\Gamma(\theta)), R_\phi \rangle^2]}$. In particular,

$$\lim_n P\left(\bigcap_{j=1}^m \sqrt{n} |h_j \circ \Gamma(\overline{X}_n) - h_j \circ \Gamma(\theta)| \le z_j \right) \ge \prod_{j=1}^m [P(\sigma_j^*|Z| \le z_j)]^{w_j}, \tag{6}$$

where w_j is the cardinality of I_j and $\sigma_j^* = \sup_{\phi \in I_j} \sigma_{\phi,j}$.

PROOF. First, let I be the set of one-to-one functions ϕ on $\{1, \dots, k\}$, with $\phi(D_h) = F_h$. From Theorem 2 in Marcheselli (2000) it follows that the regular quasi-differential τ of g is given by

$$\tau(x) = \sum_{\phi \in I} H(x_\phi) I_{A_\phi}(x),$$

where H is the differential of h at $\Gamma(\theta)$ and A_ϕ is a suitable convex set whose interior part is of type $\bigcap_{h=1}^s \{x \in \mathbb{R}^k : x_{\phi(d_{h-1}+1)} > \cdots > x_{\phi(d_h)}\}$. Since $\mathrm{Var}[\pi_j \circ H(R_\phi)] = E[\langle \nabla h_j(\Gamma(\theta)), R_\phi \rangle^2]$, from Proposition 1 the corollary immediately follows. $\square$

REMARK 4. It is easy to see that when θ has distinct components, that is, $s = k$, the function g is differentiable at θ, the cardinalities w_j are equal to 1 and inequality (6) coincides with the Šidák inequality. Moreover, in Marcheselli and Pratelli (2002) only particular cases of Corollary 2 are presented.

The relations (5) and (6) are a useful tool in constructing asymptotically conservative confidence sets for $g(\theta)$, especially when $g = h \circ \Gamma$ is not differentiable at θ. In order to show this, first let us suppose that the partition $F_1, \dots, F_s$ of $\{1, \dots, k\}$ (but not the values $\vartheta_1, \dots, \vartheta_s$ of the components of θ) is known. In this case, if $\widetilde{\sigma}^2_{\phi,j}$ denotes a consistent estimator of $\sigma^2_{\phi,j}$, owing to Corollary 2, it asymptotically turns out that

$$P\left(\bigcap_{j=1}^{m} \sqrt{n}|h_j \circ \Gamma(\overline{X}_n) - h_j \circ \Gamma(\theta)| \le z_j\right) \ge \prod_{j=1}^{m} \prod_{\phi \in I_j} \Phi(z_j/\widetilde{\sigma}_{\phi,j}) \qquad \text{a.s.,}$$

where Φ is the distribution function of $|\mathcal{N}(0,1)|$ [i.e., $\Phi(z) = \sqrt{2/\pi} \times \int_{[0,z]} e^{-x^2/2}\,dx$]. If $\widetilde{\Sigma}_\phi$ denotes the sample variance-covariance matrix obtained from the sample $X_1, \dots, X_n$ when the components of X_i are ranked according to ϕ, the random variable $\nabla h_j(\Gamma(\overline{X}_n))^t \widetilde{\Sigma}_\phi \nabla h_j(\Gamma(\overline{X}_n))$ may be considered as an estimator $\widetilde{\sigma}^2_{\phi,j}$. Alternatively, when h is a $\mathcal{C}^1$-function, n times the jackknife variance estimator for $h(\overline{X}_{n,\phi})$ may be used as an estimator $\widetilde{\sigma}^2_{\phi,j}$, where $\overline{X}_{n,\phi}$ is the sample mean of the random vectors X_i, whose components are ranked according to ϕ.

Then, even if in general the jackknife method breaks down when applied directly for the estimator $g(\overline{X}_n)$, this method may still be used for the estimator $h(\overline{X}_{n,\phi})$ in order to estimate $\sigma^2_{\phi,j}$, which through (5) makes it possible to determine asymptotically conservative confidence sets for $g(\theta)$. Finally, if the partition $F_1, \dots, F_s$ is only partially known, it suffices to apply the previous argument to a partition which includes $F_1, \dots, F_s$ [e.g., the partition given by $s = 1$ and $F_1 = \{1, \dots, k\}$]. Obviously, in a similar way, the previous procedure may be also used for any function g which is (continuously) regularly quasi-differentiable at θ.

6. An application to intrinsic diversity profile. A number of measures have been proposed to quantify the diversity of ecological populations. However, it is well known that a single diversity index is not suitable for comparing communities since the use of different indices may lead to different community rankings [see Patil and Taillie (1982)]. In order to compare ecological communities, it is convenient to consider diversity profiles [Gove, Patil, Swindel and Taillie (1994)], that is, curves depicting the simultaneous values of a large collection of diversity indices. The intrinsic diversity profile, owing to its properties, is perhaps the most important diversity profile. It is defined as the plotting of pairs (j, T_j), where

$$T_j = \frac{1}{N_1 + \cdots + N_k} \sum_{i=j+1}^{k} N_{(i)}, \qquad j = 0, \dots, k,$$

is the right-tail sum diversity index, k is the number of species, N_i is the abundance of species i and $(N_{(1)}, \dots, N_{(k)})$ is the ranked abundance vector, with $N_{(1)} \ge \cdots \ge N_{(k)}$. (The term "species" is simply a convenient label for a finite set of distinct categories comprising the community.) Thanks to intrinsic diversity profiles, it can be stated that a population C is intrinsically more diverse than C' by defining a population C to be more diverse than C' if C' leads to C through a finite sequence of (i) abundance transfer from more to less abundant species without reversing the rank-order of the species, (ii) abundance transfer to a new species and (iii) re-labelling of the species. As a matter of fact, according to Patil and Taillie (1982), the population C is intrinsically more diverse than C' if and only if the intrinsic diversity profile of C is above that of C'. Moreover, if the two profiles cross one or more times, no intrinsic ordering of the compared populations is possible. Now let $\theta = (N_1, \dots, N_k)$ be the (unknown) abundance vector of a population. Obviously $T_0 = 1, T_k = 0$ and $T_j = h_j \circ \Gamma(\theta)$, for $j = 1, \dots, k-1$, where h_j is the $\mathcal{C}^\infty$-function

$$(x_1, \dots, x_k) \mapsto \frac{1}{x_1 + \cdots + x_k} \sum_{i=j+1}^{k} x_i$$

and Γ is the decreasing ordering function on $\mathbb{R}^k$. If $\widehat{N}_1, \dots, \widehat{N}_n$ are n i.i.d. unbiased estimators for θ, with $E[\|\widehat{N}_1\|^2] < \infty$, which are frequently obtained by replicating an encounter sampling design [see Barabesi and Fattorini (1998)], let us say that $\overline{N}_n = (\widehat{N}_1 + \cdots + \widehat{N}_n)/n$ and consider

$$\widehat{T}_n = (\widehat{T}_{n,1}, \dots, \widehat{T}_{n,k-1}),$$

where $\widehat{T}_{n,j} = g_j(\overline{N}_n)$ and $g_j = h_j \circ \Gamma$. It is not difficult to note that $\widehat{T}_n$ is an asymptotically unbiased estimator for $T = (T_1, \dots, T_{k-1})$ and it turns out that $\widehat{T}_n = g(\overline{N}_n)$, $T = g(\theta)$, where $g = (g_1, \dots, g_{k-1})$ is not necessarily a Gâteaux differentiable function at θ. More precisely, g is always a twice continuously regularly quasi-differentiable function at θ but g is not Gâteaux differentiable when θ has at least two equal components.

In order to determine the asymptotic behavior of $\sqrt{n}(\widehat{T}_n - T)$ and, consequently, of $\sqrt{n}(J_n - T)$, where J_n denotes the Quenouille–Tukey jackknife estimator for $\widehat{T}_n$, let $\{n_1, \dots, n_s\}$ be the set of distinct components of θ, with $n_1 > \cdots > n_s$. Obviously $s \le k$ and $s = k$ only if all components of θ are distinct. Moreover, with the same notation as in Section 5, let $F_1, \dots, F_s$ denote the disjoint subsets of $\{1, \dots, k\}$ so that $F_h = \{i : N_i = n_h\}$ and let d_h be the cardinality of F_h, with $h = 1, \dots, s$. Finally, let $\widetilde{\Gamma} = (\widetilde{\Gamma}_1, \dots, \widetilde{\Gamma}_k) : \mathbb{R}^k \to \mathbb{R}^k$ be the Lipschitz function such that $(\widetilde{\Gamma}_l(x))_{l \in D_h}$ is the vector of the coordinates $\{x_i\}_{i \in F_h}$ ranked in decreasing order, where $D_h =]d_1 + \cdots + d_{h-1}, d_1 + \cdots + d_h] \cap \{1, \dots, k\}$. The following result is proven.

THEOREM 3. *Let R be a symmetric Gaussian random variable with the same variance-covariance matrix as $\widehat{N}_1$. Then $\sqrt{n}(\widehat{T}_n - T)$ converges in distribution to $U = (U_1, \dots, U_{k-1})$, where*

$$U_j = \frac{1}{N_1 + \cdots + N_k} \left[-T_j \sum_{l=1}^{j} \widetilde{\Gamma}_l(R) + (1 - T_j) \sum_{l=j+1}^{k} \widetilde{\Gamma}_l(R) \right].$$

If R does not have two equal components a.e., then $\sqrt{n}(J_n - \widehat{T}_n)$ converges in probability to 0 and, consequently, $\sqrt{n}(J_n - T)$ converges in distribution to U. Moreover, if $v_{n,j}$ is the jackknife variance estimator for $\widehat{T}_{n,j}$, $nv_{n,j}$ converges in probability if and only if there exists a constant c_j such that

$$\operatorname{Var}\left[-T_j \sum_{l=1}^{j} R_{\phi(l)} + (1-T_j) \sum_{l=j+1}^{k} R_{\phi(l)}\right] = c_j, \tag{7}$$

for any one-to-one function ϕ on $\{1,\dots,k\}$, with $\phi(D_h) = F_h$, $h = 1,\dots,s$. In particular, condition (7) implies that $nv_{n,j}$ converges in probability to c_j but c_j is not necessarily $E[U_j^2]$ (or $\operatorname{Var}[U_j]$).

PROOF. From Theorem 2 in Marcheselli (2000), it follows that g is a twice continuously regularly quasi-differentiable function and its regular quasi-differential at θ is $H \circ \widetilde{\Gamma}$, where H is the differential of h at $\Gamma(\theta)$. Owing to the generalization of the delta method, $\sqrt{n}(\widehat{T}_n - T)$ converges in distribution to $H \circ \widetilde{\Gamma}(R)$ which is equal to U since

$$\nabla h_j\bigl(\Gamma(\theta)\bigr) = \frac{1}{N_1 + \cdots + N_k}\left(-T_j \sum_{l=1}^{j} e_l + (1-T_j) \sum_{l=j+1}^{k} e_l\right),$$

where $(e_l)_{1\le l\le k}$ is the canonical basis of $\mathbb{R}^k$. Now let $(A_i)_i$ be a partition of $\mathbb{R}^k$ so that A_i is a convex set with interior part of the form $\bigcap_{h=1}^{s}\{x_{\sigma(f_1)} > \cdots > x_{\sigma(f_{d_h})}\}$, where $F_h = \{f_1,\dots,f_{d_h}\}$ and σ is a one-to-one function on $\{1,\dots,k\}$, with $\sigma(F_h) = F_h$. It is not difficult to observe that $(A_i)_i$ is a regular partition of $\mathbb{R}^k$ which satisfies $(*)$ and (a), (b) of Definition 2. Therefore, if R does not have two equal components a.e., Theorems 1 and 2 can be applied to g and, from the relation

$$\begin{aligned}\operatorname{Var}[L_i^\theta(\widehat{N}_1)] &= \operatorname{Var}[L_i^\theta(R)]\\ &= \frac{1}{(N_1+\cdots+N_k)^2}\operatorname{Var}\left[-T_j \sum_{l=1}^{j} R_{\phi_i(l)} + (1-T_j) \sum_{l=j+1}^{k} R_{\phi_i(l)}\right],\end{aligned}$$

where ϕ_i is a suitable one-to-one function on $\{1,\dots,k\}$, with $\phi_i(D_h) = F_h$, the theorem is proven. □

It is easy to see that U is Gaussian only if all components of θ are distinct while U_j is always a Gaussian random variable when $j = d_1 + \cdots + d_h$ and $h = 1,\dots,s$. In particular, g_j is a differentiable function at θ if and only if $j = d_1 + \cdots + d_h$ for some $h = 1,\dots,s$. Since g is a Lipschitz function at θ, the bias $E[(\widehat{T}_n - T)]$ is equivalent to $E[U]/\sqrt{n}$ and U is not generally a centered random vector. Moreover, the bias of jackknife estimator J_n is not reduced with respect to $E[U]/\sqrt{n}$. Finally, $E[(\widehat{T}_{n,j} - T_j)^2]$ is equivalent to $E[U_j^2]$ and, when

U_j is not Gaussian, the distribution of U_j is not determined by $nv_{n,j}$ even if $nv_{n,j}$ converges in probability to $E[U_j^2]$. Consequently, in order to construct an asymptotically conservative confidence band for $\widehat{T}_n - T$, the jackknife variance estimator $v_{n,j}$ cannot be used directly when θ has at least two equal components. Nevertheless, owing to Corollary 2, a confidence band for T may be obtained. As a matter of fact, the following asymptotic result can now be proven.

THEOREM 4. *For any $h = 1, \dots, s$, let I_j be a set of one-to-one functions ϕ onto $\{1, \dots, k\}$, with $\phi(D_h) = F_h$, which is an index set for the distinct functions of type $x \mapsto \sum_{l=1}^{j} x_{\phi(l)}$. For any element $(z_1, \dots, z_m) \in \mathbb{R}_+^{k-1}$, it turns out that*

$$\lim_n P\left(\bigcap_{j=1}^{k-1} \sqrt{n}|\widehat{T}_{n,j} - T_j| \le z_j\right) \ge \prod_{j=1}^{k-1} \prod_{\phi \in I_j} \Phi(z_j/\sigma_{\phi,j}), \tag{8}$$

where

$$\sigma_{\phi,j}^2 = \frac{1}{(N_1 + \cdots + N_k)^2} \operatorname{Var}\left[-T_j \sum_{l=1}^{j} R_{\phi(l)} + (1 - T_j) \sum_{l=j+1}^{k} R_{\phi(l)}\right]$$

and Φ is the distribution function of $|\mathcal{N}(0, 1)|$. In particular, if w_j is the cardinality of I_j, it holds that

$$\begin{aligned}\lim_n P\left(\bigcap_{j=1}^{k-1} \sqrt{n}|\widehat{T}_{n,j} - T_j| \le z_j\right) &\ge \prod_{j=1}^{k-1} [\Phi(z_j/\sigma_j^*)]^{w_j} \\ &\ge \prod_{h=1}^{s} [\Phi(z_j/\sigma^{*,h})]^{2^{d_h}-1},\end{aligned}$$

with $\sigma_j^* = \sup_{\phi \in I_j} \sigma_{\phi,j}$ and $\sigma^{*,h} = \sup_{j \in D_h \setminus \{k\}} \sigma_j^*$.

PROOF. Since $\nabla h_j(\Gamma(\theta)) = \frac{1}{N_1 + \cdots + N_k}(-T_j \sum_{l=1}^{j} e_l + (1 - T_j) \sum_{l=j+1}^{k} e_l)$, owing to Corollary 2, it suffices to show that

$$\prod_{j=1}^{k-1} [\Phi(z_j/\sigma_j^*)]^{w_j} \ge \prod_{h=1}^{s} [\Phi(z_j/\sigma^{*,h})]^{2^{d_h}-1}.$$

In other words, it suffices to prove that $\sum_{j \in D_h \setminus \{k\}} w_j \le 2^{d_h} - 1$. Since $w_j = \binom{d_h}{j - (d_1 + \cdots + d_{h-1})}$ for $j \in D_h$, the proof is thus complete. $\square$

Therefore, if $\tilde{\sigma}_{\phi,j}^2$ is a consistent estimator of $\sigma_{\phi,j}^2$, according to (8) an asymptotically $100(1-\alpha)\%$ conservative confidence band is given by

$$\bigcap_{h=1}^{s}\bigcap_{j\in D_h\setminus\{k\}}\{\sqrt{n}|\widehat{T}_{n,j}-T_j|\le m_{\alpha,d}\widetilde{\sigma}^{*,h}\},\tag{9}$$

where $\widetilde{\sigma}^{*,h}=\sup_{j\in D_h\setminus\{k\}}\sup_{\phi\in I_j}\widetilde{\sigma}_{\phi,j}$, $d=\sum_{j=1}^{k-1}w_j=\sum_{h=1}^{s}(2^{d_h}-1)-1$ and $m_{\alpha,d}$ is the upper α-point of the maximum of the absolute value of d independent standard Gaussian random variables.

If $\widetilde{T}_{n,j}$ denotes the vector $\widehat{T}_{n,j}\sum_{l=1}^{j}e_j+(1-\widehat{T}_{n,j})\sum_{l=j+1}^{k}e_j$ and $\widehat{N}_{1,\phi}$, $\dots,\widehat{N}_{n,\phi}$ denote the random vectors whose jth components are given, respectively, by $\pi_{\phi(j)}\circ\widehat{N}_1,\dots,\pi_{\phi(j)}\circ\widehat{N}_n$, with $j=1,\dots,k$, two consistent estimators $\widetilde{\sigma}^2_{\phi,j}$ of $\sigma^2_{\phi,j}$ may be obtained by considering

$$\frac{1}{(\pi_1(\overline{N}_n)+\cdots+\pi_k(\overline{N}_n))^2}\widetilde{T}'_{n,j}\widetilde{\Sigma}_\phi\widetilde{T}_{n,j},$$

where $\widetilde{\Sigma}_\phi$ is the sample variance-covariance matrix relative to $\widehat{N}_{1,\phi},\dots,\widehat{N}_{n,\phi}$, or by considering n times the jackknife variance estimator for $h_j(\overline{N}_{n,\phi})$, with $\overline{N}_{n,\phi}=(\widehat{N}_{1,\phi}+\cdots+\widehat{N}_{n,\phi})/n$.

In order to assess the finite sample behavior of (9), when $\widetilde{\sigma}^2_{\phi,j}$ are estimated by using jackknife variance estimators of $h_j(\overline{N}_{n,\phi})$, and to establish the behavior of the jackknife variance estimator for $\widehat{T}_n$, a Monte Carlo study was carried out on an ecological population of $k=10$ species with abundance vector $\theta=(170,40,40,35,145,10,8,6,6,6)$. Hence, a set of $B=200$ simulations was performed and for each simulation, n abundance vectors were generated as independent realizations of uniform discrete variables on $\{N_j-r_j/2+m:m=0,\dots,r_j\}$, where $(r_1,\dots,r_{10})=(30,30,70,10,30,6,4,8,4,12)$.

First, the study focused on the Monte Carlo distribution of the jackknife variance estimator $nv_{n,j}$ for $\widehat{T}_{n,j}$. Computing $nv_{n,j}$ by means of the simulation of n abundance vectors, the empirical distribution function corresponding to the B simulated realizations of $nv_{n,j}$ was obtained. Accordingly, the empirical distribution functions corresponding to $nv_{n,2},nv_{n,3},nv_{n,8}$ and $nv_{n,9}$ are displayed in Figure 1 when $n=50$. From the figure, it is at once apparent that, while the empirical distribution function corresponding to $nv_{n,2}$ resembles the distribution function of a normal random variable (which degenerates into a point for $n\to\infty$), the empirical distribution functions corresponding to $nv_{n,3},nv_{n,8}$ and $nv_{n,9}$ show a clear departure from normality (and degenerate into two points or three points for $n\to\infty$). Therefore, these results are in complete accordance with Theorem 3, even if the simulation is performed with a relatively small sample size such as $n=50$, since

$$F_1=\{1\},\qquad F_2=\{5\},\qquad F_3=\{2,3\},\qquad F_4=\{4\},$$
$$F_5=\{6\},\qquad F_6=\{7\},\qquad F_7=\{8,9,10\},$$

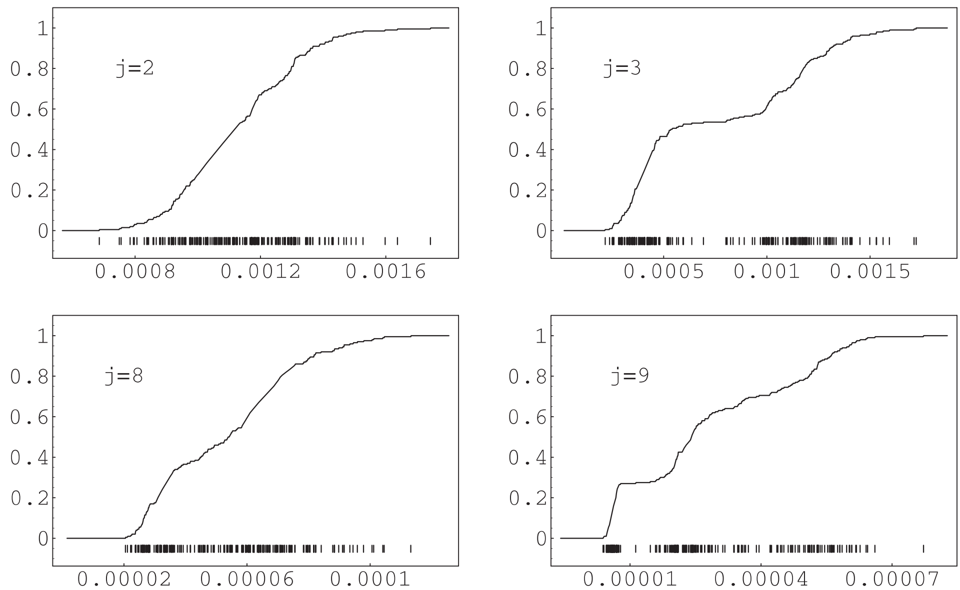

FIG. 1. *Empirical d.f.s corresponding to $nv_{n,j}$ with $n = 50$ and $j = 2, 3, 8, 9$.*

$(d_1, \dots, d_7) = (1, 1, 2, 1, 1, 1, 3)$ and from Theorem 3 it holds that the jackknife variance estimator $nv_{n,j}$ for $\widehat{T}_{n,j}$ works if and only if $j \neq 3, 8, 9$. (For the sake of simplicity the results for $nv_{n,1}, nv_{n,4}, nv_{n,5}$, $nv_{n,6}, nv_{n,7}$ are not reported in Figure 1, since they display the same behavior as $nv_{n,2}$.)

On the other hand, we analyzed the coverage of the proposed confidence band. In this case, $B = 1000$ simulations were performed for each sample size $n = 20, 50, 100, 200$. In turn, the n abundance vectors were produced by means of the previous stochastic generation. Moreover, for each simulation, the confidence bands were computed for the confidence levels $90\%, 95\%, 97.5\%$ and their coverage was assessed. The simulated coverage is ultimately reported in Table 1. These results show the excellent performance of the proposed confidence band with values very similar to the corresponding nominal levels also for relatively small sample sizes.

TABLE 1
Simulated coverage of the proposed confidence band

Level	***n***	**20**	**50**	**100**	**200**
0.900		0.93	0.94	0.95	0.95
0.950		0.96	0.97	0.98	0.98
0.975		0.98	0.98	0.99	0.99

Acknowledgment. The author is grateful to Professor L. Barabesi for his vital contribution to the realization of the simulation study.

REFERENCES

AVERSEN, J. N. (1969). Jackknifing U-statistics. *Ann. Math. Statist.* **40** 2076–2100. (Corrigendum **41** 1375.)

BARABESI, L. and FATTORINI, L. (1998). The use of replicated plot, line and point sampling for estimating species abundance and ecological diversity. *Environ. Ecol. Statist.* **5** 353–370.

GHOSH, M. (1985). Berry–Essen bounds for functionals of U-statistics. *Sankhyā Ser. A* **47** 255–270.

GOVE, J. H., PATIL, G. P., SWINDEL, B. F. and TAILLIE, C. (1994). Ecological diversity and forest management. In *Handbook of Statistics* **12**. *Environmental Statistics* (G. P. Patil and C. R. Rao, eds.) 409–462. North-Holland, Amsterdam.

MARCHESELLI, M. (1999). Some singularities in jackknifing non-smooth functions of the sample mean vector. *Metron* **57** 127–145.

MARCHESELLI, M. (2000). A generalized delta method with applications to intrinsic diversity profiles. *J. Appl. Probab.* **37** 504–510.

MARCHESELLI, M. and PRATELLI, L. (2002). Conservative confidence bands for intrinsic diversity profiles. *XLI Riunione Scientifica della Soc. Ital. Statist.* 459–462.

MILLER, R. G. (1964). A trustworthy jackknife. *Ann. Math. Statist.* **35** 1594–1605.

MILLER, R. G. (1968). Jackknifing variances. *Ann. Math. Statist.* **39** 567–582.

MILLER, R. G. (1974). An unbalanced jackknife. *Ann. Statist.* **2** 880–891.

PARR, W. C. (1985). Jackknifing differentiable statistical functionals. *J. Roy. Statist. Soc. Ser. B* **47** 56–66.

PARR, W. C. and SCHUCANY, W. R. (1982). Jackknifing L-statistics with smooth weight functions. *J. Amer. Statist. Assoc.* **77** 629–638.

PATIL, G. P. and TAILLIE, C. (1982). Diversity as a concept and its measurement (with discussion). *J. Amer. Statist. Assoc.* **77** 548–567.

REEDS, J. A. (1978). Jackknifing maximum likelihood estimates. *Ann. Statist.* **6** 727–739.

RICHMOND, J. (1982). A general method for constructing simultaneous confidence intervals. *J. Amer. Statist. Assoc.* **77** 455–460.

SHAO, J. (1993). Differentiability of statistical functionals and consistency of the jackknife. *Ann. Statist.* **21** 61–75.

SHAO, J. and TU, D. (1995). *The Jackknife and Bootstrap*. Springer, New York.

SHAO, J. and WU, C. F. J. (1989). A general theory for jackknife variance estimation. *Ann. Statist.* **17** 1176–1197.

ŠIDÁK, Z. (1967). Rectangular confidence regions for the means of multivariate normal distributions. *J. Amer. Statist. Assoc.* **62** 626–633.

THORBURN, D. E. (1977). On the asymptotic normality of the jackknife. *Scand. J. Statist.* **4** 113–118.

TUKEY, J. (1958). Bias and confidence in not-quite large samples. *Ann. Math. Statist.* **29** 614.

VARADHAN, S. R. S. (1962). Limit theorems for sums of independent random variables with values in a Hilbert space. *Sankhyā Ser. A* **24** 213–238.

DIPARTIMENTO DI METODI QUANTITATIVI
UNIVERSITÀ DI SIENA
PIAZZA SAN FRANCESCO, 8
53100 SIENA
ITALY
E-MAIL: marcheselli@unisi.it