Asymptotic equivalence of the jackknife and infinitesimal jackknife variance estimators for some smooth statistics

Alex D. Gottlieb

Abstract

The jackknife variance estimator and the the infinitesimal jackknife variance estimator are shown to be asymptotically equivalent if the functional of interest is a smooth function of the mean or a trimmed L-statistic with Hölder continuous weight function.

1 Introduction

This note concerns the asymptotic behavior of the jackknife variance estimator v_{jack}, especially regarding its relationship to the infinitesimal jackknife variance estimator v_{ijack}. We consider, in particular, the variance estimates v_{jack} and v_{ijack} for smoothly trimmed L-statistics and for smooth functions of the sample mean. We prove that v_{jack} and v_{ijack} are often asymptotically equivalent to one another in the sense that

$$v_{jack} - v_{ijack} \;=\; \mathrm{O_p}(n^{-h}) \tag{1.1}$$

for some $h > 0$. The equivalence of v_{jack} and v_{ijack} can sometimes be used to prove that v_{jack} is asymptotically normal, but — remarkably — it also holds even when these estimators are not asymptotically normal.

These issues are relevant to the following scenario of statistical practice: One wishes to estimate some functional $T(p)$ of an unknown population distribution p, and, to this end, one draws n samples from the population and uses $T(\epsilon_n)$ to estimate $T(p)$, where ϵ_n is the empirical distribution of the n samples. One would then like an estimate of the sampling variance of $T(\epsilon_n)$. Two widely-used nonparametric estimates of this variance are v_{jack} and v_{boot}, the jackknife and bootstrap variance estimates. Having obtained one of these estimates, one naturally desires to know how accurate it is. To assess the accuracy of the usual Monte Carlo approximation of v_{boot}, one may use the jackknife-after-bootstrap technique of Efron (1992). This note concerns the asymptotic behavior of v_{jack} and a closely related estimator v_{ijack}, the infinitesimal jackknife.

Beran(1984) showed that v_{jack}, v_{ijack}, and v_{boot} are asymptotically equivalent, and asymptotically normal, if the functional $T(p)$ has a well-behaved second-order functional derivative.

While the proof of Beran (1984) requires a strong statement of the Dvoretsky-Kiefer-Wolfowitz inequality to handle the asymptotics of v_{boot}, the asymptotic equivalence of v_{jack} and v_{ijack} is easier to prove, and does not require second-order differentiability of T. Indeed, we shall see that v_{jack} and v_{ijack} are asymptotically equivalent if T has a well-behaved first-order derivative and v_{ijack} is consistent as an estimator of the variance of $T(\epsilon_n)$.

The equivalence of v_{jack} to v_{ijack} can help one to determine the asymptotic variance of the former. For example, Gardiner and Sen (1979) have carefully studied the asymptotic normality of v_{ijack} in the context of L-statistics. We shall see in Section 4 that their work establishes the asymptotic normality of v_{jack}, too, thanks to the equivalence of v_{jack} and v_{ijack} for variance estimation of L-statistics.

Sometimes v_{jack} and v_{ijack} are asymptotically equivalent even when they are not asymptotically normal. Consider, for example, the estimation of the variance of a function g of the sample mean. If g is once, but not twice, continuously differentiable, then v_{jack} and v_{ijack} may not be asymptotically normal, yet they will satisfy (1.1) as long as g' is Hölder continuous of order h.

After the necessary definitions are presented in the next section, we prove in Section 3 that v_{jack} and v_{ijack} are asymptotically equivalent as estimators of the variance of smooth functions of the sample mean. In Section 4 we discuss the asymptotic normality of v_{jack} as an estimator of the variance of trimmed L-statistics.

2 Background and definitions

Let p be a probability measure on a sample space $\mathcal{X}$. Given n samples from $\mathcal{X}$, sampled independently under the probability law p, one desires to estimate the value $T(p)$ of some real functional T on the space $\mathcal{P}(\mathcal{X})$ of all probability measures on $\mathcal{X}$. Denote by ϵ_n the map that converts n data points $x_1, x_2, \ldots, x_n$ into the empirical measure

$$\epsilon_n(x_1, x_2, \ldots, x_n) \;=\; \frac{1}{n}\sum_{i=1}^{n} \delta(x_i) \tag{2.1}$$

where $\delta(x_i)$ denotes a point-mass at x_i. The *plug-in estimate* of $T(p)$ given the data $\mathbf{x} = (x_1, \ldots, x_n)$ is

$$T_n \;=\; T(\epsilon_n(\mathbf{x})). \tag{2.2}$$

Suppose T_n is an asymptotically normal estimator of $T(p)$, i.e., suppose the distribution of $n^{1/2}(T_n - T(p))$ tends to $\mathcal{N}(0, \sigma^2)$. The jackknife is a computational technique for estimating σ^2: one transforms the n original data points into n *pseudovalues* and computes the sample variance of those pseudovalues.

Given the data $\mathbf{x} = x_1, x_2, \ldots, x_n$, the *jackknife pseudovalues* are

$$Q_{ni} \;=\; nT_n(\epsilon_n) \;-\; (n-1)T(\epsilon_{ni}) \qquad\qquad i = 1, 2, \ldots, n$$

with ϵ_n as in (2.1) and

$$\epsilon_{ni} \;=\; \frac{1}{n-1}\sum_{j\neq i}\delta(x_j). \tag{2.3}$$

The *jackknife variance estimator* is

$$v_{jack}(x_1, x_2, \ldots, x_n) \;=\; \frac{1}{n-1}\sum_{i=1}^{n}\left(Q_{ni} - \overline{Q_n}\;\right)^2 \tag{2.4}$$

where $\overline{Q_n} = \frac{1}{n}\sum Q_{nj}$. The variance estimator v_{jack} is said to be *consistent* if $v_{jack} \longrightarrow \sigma^2$ almost surely as $n \to \infty$. Sufficient conditions for the consistency of v_{jack} are given in terms of the functional differentiability of T. An early result of this kind states that v_{jack} is consistent if T is strongly Fréchet differentiable (Parr(1985)), and it is now known that v_{jack} is consistent even if T is only continuously Gâteaux differentiable as defined in Shao(1993).

A functional derivative of T at p, denoted ∂T_p, is a linear functional that best approximates the behavior of T near p in some sense. For instance, a functional T on the space of bounded signed measures $\mathcal{M}(\mathcal{X})$ is *Gâteaux differentiable* at p if there exists a continuous linear functional ∂T_p on $\mathcal{M}(X)$ such that

$$\lim_{t\to 0}\left|t^{-1}\left(T(p+tm) - T(p)\right) \;-\; \partial T_p(m)\right| \;=\; 0$$

for all $m \in \mathcal{M}(\mathcal{X})$. The concept of Hadamard differentiability is more relevant to statistical asymptotics, for the fluctuations of $T(\epsilon_n)$ about $T(p)$ are asymptotically normal if T is Hadamard differentiable at p. A functional $T : \mathcal{P}(\mathbb{R}) \longrightarrow \mathbb{R}$ is *Hadamard differentiable* at p if there exists a continuous linear functional ∂T_p on $\mathcal{M}(\mathbb{R})$ such that

$$\lim_{t\to 0}\left|t^{-1}\left(T(p+tm_t) - T(p)\right) \;-\; \partial T_p(m)\right| \;=\; 0$$

whenever $\{m_t\}_{t\in\mathbb{R}}$ is such that $\lim_{t\to 0} m_t = m$ and $m_t(\mathbb{R}) = 0$ for all t, the topology on $\mathcal{M}(\mathbb{R})$ being the one induced by the norm $\|m\| = \sup_{t\in\mathbb{R}}\{|m((-\infty, t])|\}$. If T is Hadamard differentiable at p, the variance of $n^{1/2}T(\epsilon_n)$ tends to

$$\sigma^2 \;=\; \mathbb{E}_p\phi_p^2 \tag{2.5}$$

as $n \longrightarrow \infty$, where $\phi_p(x)$ is the *influence function*

$$\phi_p(x) \;=\; \partial T_p(\delta(x) - p) \tag{2.6}$$

(this can be shown via the Delta Method using Donsker's theorem (van der Waart(1998))). The *infinitesimal jackknife estimator* (Jaeckel(1972)) of σ^2 is obtained by substituting the empirical measure ϵ_n for p in (2.5):

$$v_{ijack} \;=\; \mathbb{E}_{\epsilon_n}\phi_{\epsilon_n}^2. \tag{2.7}$$

3 Functions of the mean

When q is a measure, we denote $\int xq(dx)$ by $\overline{q}$ if the integral is defined. Let $g \in C^1(\mathbb{R})$ and let

$$T(m) \;=\; g\left(\overline{m}\right)$$

be defined for all finite signed measures m with finite first moment. The functional derivative at m of T, evaluated at q, is $\partial T_m(q) \;=\; g'\left(\overline{m}\right)\overline{q}$; the influence function (2.6) is $\phi_m(x) \;=\; g'\left(\overline{m}\right)\left(x - \overline{m}\right)$. Suppose that $x_1, x_2, \ldots$ are iid p, and p has a finite second moment. Let T_n denote the plug-in estimator defined in (2.2). Then the asymptotic variance of $n^{1/2}\left(T_n - T(p)\right)$ is

$$\sigma^2 \;=\; g'(\overline{p})^2\left\{\int x^2 p(dx) - \overline{p}^2\right\}. \tag{3.1}$$

Let v_{jack} and v_{ijack} denote the jackknife and infinitesimal jackknife variance estimates of σ^2.

Proposition 3.1 *If g' is Hölder continuous of order $h > 1/2$ (with global Hölder constant) and p has a finite moment of order $2 + 2h$ then*

$$v_{jack} - v_{ijack} \;=\; \mathrm{O_p}(n^{-h}).$$

Proof: Setting $\Delta_{ni} = \left(Q_{nj} - \overline{Q_n}\right) - \phi_{\epsilon_n}(x_i)$, formula (2.4) for v_{jack} yields

$$v_{jack} \;=\; \mathbb{E}_{\epsilon_n}\phi^2_{\epsilon_n} \;+\; \frac{1}{n-1}\mathbb{E}_{\epsilon_n}\phi^2_{\epsilon_n} \;+\; \frac{2}{n-1}\sum_{i=1}^{n}\phi_{\epsilon_n}(x_i)\Delta_{ni} \;+\; \frac{1}{n-1}\sum_{i=1}^{n}\Delta^2_{ni}\,. \tag{3.2}$$

The second term on the right hand side of (3.2) is $\mathrm{O_p}\left(1/n\right)$ since

$$\mathbb{E}_{\epsilon_n}\phi^2_{\epsilon_n} \;=\; \frac{1}{n}\sum_{i=1}^{n}\phi^2_{\epsilon_n}(x_i) \;=\; \frac{1}{n}\sum_{i=1}^{n} g'\left(\overline{\epsilon_n}\right)^2\left(x_i - \overline{\epsilon_n}\right)^2$$

converges almost surely to σ^2.

To control the last two terms on the right hand side of (3.2) we need a bound on Δ_{ni}. Recall the notation ϵ_{ni} of (2.3). Since g is differentiable, $g\left(\overline{\epsilon_{nj}}\right) - g\left(\overline{\epsilon_{ni}}\right) \;=\; g'\left(\eta_{ji}\right)\left(\overline{\epsilon_{nj}} - \overline{\epsilon_{ni}}\right)$ for some η_{ji} between $\overline{\epsilon_{ni}}$ and $\overline{\epsilon_{nj}}$, so that

$$Q_{ni} - \overline{Q_n} \;=\; \frac{n-1}{n}\sum_{j=1}^{n}\left(g\left(\overline{\epsilon_{nj}}\right) - g\left(\overline{\epsilon_{ni}}\right)\right) \;=\; \frac{n-1}{n}\sum_{j=1}^{n} g'\left(\eta_{ji}\right)\left(\overline{\epsilon_{nj}} - \overline{\epsilon_{ni}}\right).$$

Therefore, since $\phi_{\epsilon_n}(x_i) = g'\left(\overline{\epsilon_n}\right)\left(x_i - \overline{\epsilon_n}\right) = \frac{1}{n}\sum_j g'\left(\overline{\epsilon_n}\right)\left(x_i - x_j\right)$,

$$\begin{aligned}\Delta_{ni} \;=\; \left(Q_{ni} - \overline{Q_n}\right) - \phi_{\epsilon_n}(x_i) \;&=\; \frac{n-1}{n}\sum_{j=1}^{n} g'\left(\eta_{ji}\right)\left(\overline{\epsilon_{nj}} - \overline{\epsilon_{ni}}\right) \;-\; \frac{1}{n}\sum_{j=1}^{n} g'\left(\overline{\epsilon_n}\right)\left(x_i - x_j\right)\\ &=\; \frac{1}{n}\sum_{j=1}^{n}\left(g'\left(\eta_{ji}\right) - g'\left(\overline{\epsilon_n}\right)\right)\left(x_i - x_j\right).\end{aligned}$$

But g' is Hölder continuous of order h and $|\eta_{ji} - \overline{\epsilon_n}| < \max\{|\overline{\epsilon_{nj}} - \overline{\epsilon_n}|, |\overline{\epsilon_{ni}} - \overline{\epsilon_n}|\}$, so the identity $(n-1)(\epsilon_n - \epsilon_{ni}) \;=\; \delta_{x_i} - \epsilon_n$ implies that

$$|g'(\eta_{ji}) - g'(\overline{\epsilon_n})| \;\le\; C(|\overline{\epsilon_{nj}} - \overline{\epsilon_n}|^h + |\overline{\epsilon_{ni}} - \overline{\epsilon_n}|^h) \;\le\; C(n-1)^{-h}(|\overline{\epsilon_n} - x_j|^h + |\overline{\epsilon_n} - x_i|^h),$$

where C is a global Hölder constant for g'. It follows that

$$|\Delta_{ni}| \;=\; C(n-1)^{-h}\frac{1}{n}\sum_{j=1}^{n}(|\overline{\epsilon_n} - x_j|^h + |\overline{\epsilon_n} - x_i|^h)(|\overline{\epsilon_n} - x_j| + |\overline{\epsilon_n} - x_i|).$$

With this bound on Δ_{ni}, and assuming that p has a finite moment of order $2(1+h)$, it may be shown that

$$\frac{1}{n}\sum_{i=1}^{n}\Delta_{ni}^2 \;=\; \mathrm{O_p}(n^{-2h}),$$

and then, by the Cauchy-Schwartz inequality, that

$$\Big|\frac{1}{n}\sum_{i=1}^{n}\phi_{\epsilon_n}(x_i)\Delta_{ni}\Big| \;=\; \mathrm{O_p}(n^{-h}).$$

Substituting the preceding estimates in (3.2) completes the proof. □

Consider $g(x) = x - \mathrm{sgn}(x)x^2$. This function has a Lipschitz continuous derivative but does not have a second-order derivative at 0. By Proposition 3.1,

$$v_{jack} - v_{ijack} \;=\; \mathrm{O_p}(1/n).$$

However, for some population distributions having mean 0, one can prove that v_{ijack} is not asymptotically normal, and simulations suggest that $v_{jack} - v_{boot}$ is $\mathrm{O_p}(1/\sqrt{n})$ rather than $\mathrm{O_p}(1/n)$. This example shows that v_{jack} is more closely related to v_{ijack} than it is to v_{boot}.

4 Trimmed L-statistics

Suppose that $\ell : (0,1) \longrightarrow \mathbb{R}$ is supported on $[\alpha, 1-\alpha]$ for some $0 < \alpha < 1/2$, and let

$$L(p) \;=\; \int_0^1 P^{-1}(s)\ell(s)ds. \tag{4.1}$$

Here P^{-1} denotes the quantile function for p, i.e., $P^{-1}(s) = \min\{x : P(x) \ge s\}$ for $0 < s < 1$ where P denotes the cdf of p. A plug-in estimate for L is called a *trimmed L-statistic*, or a trimmed *linear combination of quantiles.*

If the weight function ℓ is continuous then L is Hadamard differentiable at all $p \in \mathcal{P}(\mathbb{R})$ (see, e.g., Lemma 22.10 of van der Waart (1998)), and so the L-statistics are asymptotically normal (an original reference is Stigler (1974)). The asymptotic variance σ^2 of the L-statistics may be estimated by v_{jack}, which converges almost surely to σ^2 if ℓ is continuous (Parr(1985), Shao and Tu (1995)). The jackknife and infinitesimal jackknife would seem to be the only nonparametric methods of consistent variance estimation for L-statistics, aside from the bootstrap (Shucany and Parr(1982)).

We turn now to the question of the asymptotic normality of v_{jack}. In this regard, a variant of the L-functional (4.1) has been treated in the literature, namely

$$\mathcal{L}(p) = \int x\ell(P(x))p(dx) \ . \tag{4.2}$$

If P is continuous and strictly increasing then $\mathcal{L}$ of (4.2) equals L of (4.1). Beran (1984) proves that v_{jack} for $\mathcal{L}$ is asymptotically normal — and so is v_{boot} — if ℓ is continuously differentiable and p has bounded support. Section 2.2.3 of Shao and Tu (1995) incorrectly claims that v_{jack} for $\mathcal{L}$ is asymptotically normal if ℓ is Hölder continuous of order greater than 1/2, and it also wrongly claims that the asymptotic variance equals $\mathrm{Var}(\phi_p^2)$, where ϕ_p is the influence function of $\mathcal{L}$. A detailed discussion of those errors is given in an unpublished technical report (Gottlieb (2001)). Nevertheless, a reworking of Definition 2.6 and Theorem 2.7 in Shao and Tu (1995) leads us to the following general proposition, which will presently be applied to the case where the $T(\epsilon_n)$ are L-statistics:

Proposition 4.1 *Let ϵ_n denote the empirical distribution of n iid samples from p, and let v_{jack} and v_{ijack} denote the jackknife and infinitesimal jackknife estimates of the variance of $T(\epsilon_n)$.*

Let $\|q' - q\|$ denote the supremum of the absolute value of the difference between the cdf's of q' and q. Suppose that there exist positive constants C, δ, and h such that

$$T(q') \ = \ T(q) + \partial_q T(q' - q) + \mathcal{R}(q', q) \tag{4.3}$$

for all q', q with $\|q' - p\|, \|q - p\| < \delta$, where the remainder $|\mathcal{R}(q', q)| \leq C\|q' - q\|^{1+h}$. Then

$$v_{jack} - v_{ijack} \ = \ \mathrm{O_p}(n^{-h})$$

if v_{ijack} is bounded in probability.

The straightforward proof of this proposition proceeds like the proof of Proposition 3.1 above, except that (4.3) is used to bound Δ_{ni} in (3.2).

Now, let $\mathcal{L}$ be a trimmed L-functional of the form (4.2) whose weight function ℓ is Hölder continuous of order h. Upon integrating the right hand side of (4.2) by parts, it becomes easy to verify that $\mathcal{L}$ admits the expansion (4.3) near any p. Since v_{ijack} converges almost surely, Proposition 4.1 implies that v_{jack} and v_{ijack} are asymptotically equivalent.

This equivalence allows us to conclude that v_{jack} is asymptotically normal for many L-functionals, for Gardiner and Sen (1979) have found hypotheses that guarantee the asymptotic normality of v_{ijack} for generalized L-functionals of the form (4.1). They begin by assuming that the cdf P of the population distribution is continuous. In this case

$$v_{ijack} \;=\; \mathbb{E}_{\epsilon_n}\phi^2_{\epsilon_n} \;=\; \int\int \ell(P_n(y))\,[P_n(y\wedge z) - P_n(y)P_n(z)]\,\ell(P_n(z))dydz\;.$$

In order to make contact with the work of Gardiner and Sen (1979), let us suppose that P is continuous and strictly increasing. Their hypotheses are general enough to apply to non-trimmed L-statistics, but too complicated to be repeated here. Suffice it to say that their theorem applies under our current assumptions that ℓ is trimmed and that P has no jumps or flats, if it is assumed in addition that P does not have very heavy tails and that ℓ is piecewise continuously differentiable with Hölder continuity of order greater than 1/2 at the cusps. (Imagine, for example, a piecewise-linear ℓ whose graph is shaped like a desert mesa; this is one of the weight functions recommended in Stigler (1973) for smoothly trimmed means.) In these cases v_{jack} is asymptotically normal as well, by Proposition 4.1.

5 Acknowledgments

The author would like to thank Steve Evans for his advice and encouragement, and also Rudolf Beran. Support from the ESI and WPI in Vienna and the START project *Nonlinear Schrödinger and quantum Boltzmann equations* (FWF Y-137) is acknowledged.

References

[1] Beran, R. (1984). Jackknife approximations to bootstrap estimates. *The Annals of Statistics*, **12** (1), 101 - 118.

[2] Efron, B. (1992). Jackknife-after-bootstrap sample errors and influence functions. *Journal of the Royal Statistical Society B*, **54**, 83 - 127.

[3] Gardiner, J. C. and Sen, P. K. (1979). Asymptotic normality of a variance estimator of a linear combination of a function of order statistics. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **50**, 205 - 221.

[4] Gottlieb, A. D. (2001). Asymptotic accuracy of the jackknife variance estimator for certain smooth statistics. *Preprint server lanl.arXiv.org* math.PR/0109002

[5] Jaeckel, L. (1972). The infinitesimal jackknife. *Bell Laboratories Memorandum* MM 72-1215-11.

[6] Parr, W. C. and Shucany, W. R. (1982). Jackknifing L-statistics with smooth weight functions. *Journal of the American Statistical Association*, **77**, 629 - 638.

[7] Parr, W. C. (1985). Jackknifing differentiable statistical functions. *Journal of the Royal Statistical Society B*, **47** (1), 56 - 66.

[8] Shao, J. (1993). Differentiability of statistical functionals and consistency of the jackknife. *The Annals of Statistics*, **21** (1), 61 - 75.

[9] Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap*. Springer-Verlag, New York.

[10] Stigler, S. M. (1973). The asymptotic distribution of the trimmed mean. *The Annals of Statistics*, **1**, 472 - 477.

[11] Stigler, S. M. (1974). Linear functions of order statistics with smooth weight functions. *The Annals of Statistics*, **2**, 676 - 693.

[12] van der Waart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.