The Annals of Statistics
1981, Vol. 9, No. 3, 694–696

ON BERRY-ESSEEN RATES FOR JACKKNIFE ESTIMATORS[1]

BY K. F. CHENG[2]

The Florida State University

Consider an ordinary estimation problem for an unknown parameter θ. Let the estimator θ_n^* be the jackknife of a function of a U-statistic. Under mild assumptions, we demonstrate that $\sup_t |P[n^{1/2}(\theta_n^* - \theta)/S_n^* \le t] - \Phi(t)| = O(n^{-p/2(p+1)})$, where S_n^{*2} is a jackknife estimator of the asymptotic variance of $n^{1/2}\theta_n^*$, $\Phi(t)$ is the standard normal distribution and p is some positive number.

1. Introduction. In the past there has been extensive research on the technique of *jackknife*. Conclusions on the limit distribution for the jackknife of a function of several means were rigorously obtained by Miller (1964), and extension to the jackknife of a function of several *U-statistics* was carried out by Arvesen (1969). Interesting examples belonging to this setting include ratios, the t-statistic, the Wilcoxon signed-rank test, and product-moment correlation coefficient. (See Gray and Schucany (1972) or Miller (1974) for reviews of the literature.) Some invariance principles relating to jackknife estimators were examined by Sen (1977). These results are useful in sequential analysis. The object of the present note is to explore the *Berry-Esseen rates* associated with asymptotic normality for a large class of jackknife statistics.

We shall be concerned with the jackknife of a function of a U-statistic. Let $h(x_1, \cdots, x_m)$, symmetric in its arguments, be a Borel-measurable *kernel*. Based on the i.i.d. random variables $X_1, \cdots, X_n (n \ge m)$, the corresponding U-statistic U_n for estimation of $\xi = Eh(X_1, \cdots, X_m)$, is formed by averaging the kernel h symmetrically over the observations. Namely,

$$U_n = U(X_1, \cdots, X_n) = \binom{n}{m}^{-1} \sum\nolimits_{C_{n,m}} h(X_{i_1}, \cdots, X_{i_m}), \tag{1}$$

where $C_{n,m}$ denotes the set of the $\binom{n}{m}$ combinations of m distinct elements $\{i_1, \cdots, i_m\}$ from $\{1, 2, \cdots, n\}$.

Let $g(x)$ be a function defined on R and $\theta = g(\xi)$ be the parameter of interest. A natural estimator of θ based on U_n is defined by

$$\hat{\theta}_n = g(U_n).$$

Thus the associated *jackknife estimator* θ_n^* is given by

$$\theta_n^* = ng(U_n) - (n-1)(n^{-1} \sum\nolimits_{i=1}^n g(U_{n-1}^i)), \tag{2}$$

where $U_{n-1}^i = U(X_1, \cdots, X_{i-1}, X_{i+1}, \cdots, X_n)$ is obtained after elimination of the ith observation. A related representation of θ_n^* follows from the *martingale* structure of U-statistics. For any $n \ge 1$, denote by $\mathbf{X}_{(n)}$ the order statistics $(X_{n1}, \cdots, X_{nn})$. Let $\mathscr{F}_n = \sigma\{\mathbf{X}_{(n)}, X_{n+1}, X_{n+2}, \cdots\}$ be the σ-field generated by $\mathbf{X}_{(n)}$ and X_i, $i \ge n+1$. If $E|h| < \infty$, then the sequence $\{U_n, \mathscr{F}_n\}_{n \ge m}$ forms a reverse martingale (see Berk (1966)). Hence, following Sen

Received January, 1980; revised May, 1980.

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

[2] Now at SUNY at Buffalo, Amherst, New York.

AMS 1970 *subject classification.* Primary 60B10; secondary 62G05.

Key words and phrases. Jackknife; U-statistic; Berry-Esseen rates.

(1977), and utilizing Taylor expansion of the function g to the second order, (2) can be represented as

$$\begin{aligned}\theta_n^* &= g(\xi) + g'(\xi)(U_n - \xi) + \tfrac{1}{2}g''(\bar{U}_n)(U_n - \xi)^2 \\ &\quad + (n-1)E\{\tfrac{1}{2}g''(\hat{U}_n)(U_{n-1} - U_n)^2 \mid \mathscr{F}_n\}, \qquad \text{a.s.}\end{aligned} \tag{3}$$

if the function g has a second derivative, where $\bar{U}_n$ lies between ξ and U_n and $\hat{U}_n$ lies between U_{n-1} and U_n. This representation for the jackknife estimator will be used in the sequel.

2. Main theorem. Let

$$h_i(x_1, \cdots, x_i) = Eh(x_1, \cdots, x_i, X_{i+1}, \cdots, X_m)$$

and

$$\zeta_i = E(h_i^2 - \xi^2), \qquad i = 0, 1, \cdots, m.$$

Then the asymptotic variance of $n^{1/2}\theta_n^*$ is $m^2\zeta_1[g'(\xi)]^2$ which can be consistently estimated by $S_n^{*2} = (n-1)\sum_{i=1}^n [g(U_{n-1}^i) - n^{-1}\sum_{i=1}^n g(U_{n-1}^i)]^2$. Our main theorem deals with the rate of convergence in the central limit theorem for $n^{1/2}(\theta_n^* - g(\xi))/S_n^*$. Let Φ be the standard normal distribution.

THEOREM. *Let $m = 2$. If $Eh^{2p}(X_1, X_2) < \infty$, $Eh^{2p}(X_1, X_3)h^{2p}(X_2, X_3) < \infty$, $E\,|\,h(X_1, X_2)^{4.5}\,| < \infty$, $\zeta_1 > 0$, $g'(\xi) \neq 0$, and g'' is bounded, then*

$$\sup_t |\,p[n^{1/2}(\theta_n^* - g(\xi))/S_n^* \le t] - \Phi(t)\,| = O(a_n) \qquad as\ n \to \infty,$$

where $a_n = n^{-p/2(p+1)}$. In particular, if $h(X_1, X_2)$ has finite moments of all orders, then $a_n = n^{-(1/2)+\varepsilon}$ for any arbitrary small $\varepsilon > 0$.

PROOF. For the case of $g'(\xi) > 0$, we define $S_n^2 = (n-1)\sum_{i=1}^n (U_{n-1}^i - U_n)^2$ and decompose $g'(\xi)S_n(S_n^*)^{-1}$ into

$$g'(\xi)S_n(S_n^*)^{-1} = 1 - V_n/2S_n^2 + O((V_n/S_n^2)^2)$$

except on a set with probability $O(n^{-1/2})$, where V_n is a quantity such that

$$P((V_n/S_n^2)^2 > n^{-1/2}\ln^{-1} n) = O(n^{-1/2}).$$

Accordingly, utilizing a result by Callaert and Veraverbeke (1980) we have

$$\begin{aligned}P(|\,(U_n - \xi)/S_n \cdot (V_n/S_n^2)^2\,| > n^{-1/2}) &\le P(|\,(U_n - \xi)/S_n\,| > \ln n) \\ &\quad + P(|\,V_n/S_n^2\,|^2 > n^{-1/2}\ln^{-1} n) = O(n^{-1/2}),\end{aligned}$$

and hence following (3),

$$\begin{aligned}&\sup_t |\,P(n^{1/2}(\theta_n^* - g(\xi))/S_n^* \le t) - \Phi(t)\,| \\ &\qquad \le \sup_t |\,P(n^{1/2}(U_n - \xi)/S_n \le t) - \Phi(t)\,| \\ &\qquad + P(|\,n^{1/2}(U_n - \xi)/S_n \cdot (V_n/2S_n^2)\,| > b_n) + O(b_n) + P(|\,\eta_n/S_n^*\,| > c_n) + O(c_n),\end{aligned}$$

where b_n and c_n are any two sequences of positive constants and $\eta_n = n^{1/2}(\theta_n^* - g(\xi) - (U_n - \xi)g'(\xi))$. Take $b_n = n^{-p/(2p+1)}(\ln n)^{2p/(2p+1)}$, $c_n = n^{-p/2(p+1)}$, and appeal to a theorem by Callaert and Veraverbeke (1980) we obtain

$$\begin{aligned}\sup_t |\,P(n^{1/2}(\theta_n^* - g(\xi))/S_n^* \le t) - \Phi(t)\,| &= O(n^{-1/2}) \\ &\quad + O(n^{-p/(2p+1)}(\ln n)^{2p/(2p+1)}) + O(n^{-p/2(p+1)}) = O(n^{-p/2(p+1)}).\end{aligned}$$

The argument for the case of $g'(\xi) < 0$ is similar. □

REMARK. The condition on g'' can be reduced to merely that g'' is bounded in a neighborhood of ξ.

Acknowledgment. The author thanks the referee and editor for their comments leading to an improvement in the presentation of this paper.

REFERENCES

ARVESEN, J. M. (1969). Jackknifing U-statistics. *Ann. Math. Statist.* **40** 2076–2100.

BERK, R. H. (1966). Limiting behavior of posterior distribution when the model is incorrect. *Ann. Math. Statist.* **37** 51–58.

CALLAERT, H. and VERAVERBEKE, N. (1980). The order of normal approximation for a studentized U-statistic. *Ann. Statist.* **9** 193–199.

GRAY, H. L. and SCHUCANY, W. R. (1972). *The Generalized Jackknife Statistic.* Dekker, New York.

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

MILLER, R. G., JR. (1974). The jackknife—A review. *Biometrika* **61** 1–15.

SEN, P. K. (1977). Some invariance principles relating to jackknifing and their role in sequential analysis. *Ann. Statist.* **5** 316–329.

DEPARTMENT OF STATISTICS
SUNY AT BUFFALO
4230 RIDGE LEA ROAD
AMHERST, NEW YORK 14226