The Annals of Statistics
1994, Vol. 22, No. 4, 2011–2030

ON GENERAL RESAMPLING ALGORITHMS AND THEIR PERFORMANCE IN DISTRIBUTION ESTIMATION

BY PETER HALL AND ENNO MAMMEN

Australian National University

Recent work of several authors has focussed on first-order properties (e.g., consistency) of general bootstrap algorithms, where the numbers of times that data values are resampled form an exchangeable sequence. In the present paper we develop second-order properties of such algorithms, in a very general setting. Performance is discussed in the context of distribution estimation, and formulae for higher-order moments and cumulants are developed. Arguing thus, necessary and sufficient conditions are given for general resampling algorithms to correctly capture second-order properties.

1. Introduction. The classical bootstrap may be thought of as a rather special device for constructing a new data sequence having the same size as the original sample. All of the members of the new sequence are drawn from the original sample, and are present in proportions which are determined by a uniform multinomial distribution on the original sample values. Of course, the latter distribution is a consequence of the "random sampling, with replacement" concept that underlies the classical bootstrap algorithm. There are several alternative versions of this scheme, some of which keep the "random sampling" part of the algorithm but remove the requirement that the elements of the resample be a subset of those in the original sample. The smoothed bootstrap is an example; see Silverman and Young (1987) and Young (1988).

Another approach is to retain the requirement that the elements of the resample be a subset of those of the original sample, but to relax the assumption of random resampling. A very significant early move in this direction was made by Rubin (1981), with his development of the Bayesian bootstrap. There, the numbers of sample values in the resample were chosen according to a Dirichlet distribution. Second-order properties of this prescription were studied by Weng (1989), who showed that it is not as accurate as the percentile bootstrap when employed to approximate the frequentist distribution of the sample mean, but that it is more accurate when used to describe a prior distribution. See also Lo (1987). Other developments in the same vein include the so-called wild bootstrap, whose origins may be traced to variance estimation in heteroscedastic regression [Wu (1986) and Beran (1986)] and which has been studied more recently in the contexts of linear statistical inference [Liu (1988)] and nonparametric curve estimation [Härdle and Mammen (1990)]. The wild bootstrap is usually explicitly designed so that it correctly reproduces second-order proper-

Received July 1992; revised January 1994.
AMS 1991 *subject classifications.* Primary 62G05; secondary 62G15.
Key words and phrases. Bootstrap, Cornish–Fisher expansions, cumulant, distribution estimation, Edgeworth expansion, jackknife, mean, moment, resample, wild bootstrap.

ties of sampling distributions. Mason and Newton (1992) have discussed first-order properties for general resampling schemes with exchangeable properties, including the schemes based on Efron's (1979) classical definition of the bootstrap and also Rubin's Bayesian-motivated scheme. Barbe and Bertrail (1993) have also treated general forms of the weighted bootstrap.

Our aim in the present paper is to examine second-order properties of a very large class of bootstrap alternates. These involve constructing a resample whose values are selected from the sample according to very general sampling schemes. The latter include both the Dirichlet scheme of the Bayesian bootstrap and that of the wild bootstrap. In this sense our work might be seen as a generalization of that of authors such as Weng (1989) and Mason and Newton (1992). However, our outlook is not motivated specifically by Bayesian considerations, and we analyse different approaches entirely from a frequentist viewpoint. We study the case of estimating the distribution of the mean, and treat both percentile and percentile-t arguments. Other contexts, such as smooth functions of multivariate means, could be treated similarly but with very much more complex notation.

In detail, our model is as follows. Suppose we are given a random sample $\mathcal{X} = \{X_1, \dots, X_n\}$, from which we draw a resample where the number of times X_i appears is given by the random variable N_{ni}^*. We assume that the N_{ni}^*'s are exchangeable. Section 2 develops expansions for cumulants and distribution functions under such very general schemes, in terms of moments of the N_{ni}^*'s. By asking that

$$\sum_{i=1}^{n} N_{ni}^* = n, \tag{1.1}$$

we ensure that the resample is of the same size as the original sample. Of our examples, only in one version of the wild bootstrap is this a restriction, and there the N_{ni}^*'s are independent. We do treat that case. In all circumstances, our expansions are compared directly to their counterparts where the N_{ni}^*'s are independent and enjoy the same marginal distribution. Quite apart from the fact that this approach enables us to include the exceptional version of the wild bootstrap, it provides a particularly convenient and informative foil for comparison in general.

Since we are restricting our attention to means then there is no real difficulty in allowing N_{ni}^* to take noninteger, or even negative, values. We ask that in such cases the mean and the mean of the squares of the resample values be defined by

$$\overline{X}_1 = n^{-1} \sum_{i=1}^{n} N_{ni}^* X_i \quad \text{and} \quad \overline{X}_2 = n^{-1} \sum_{i=1}^{n} N_{ni}^* X_i^2,$$

which avoids any ambiguity. Example 2 in Section 2 is of this form.

Our results immediately yield necessary and sufficient conditions for second-order accuracy of various generalised bootstrap methods, based on both percentile and percentile-t arguments. For example, the classical percentile and percentile-t bootstrap techniques are second-order accurate, in the sense of correctly capturing skewness-based departures from the sampling distribution of the mean, if and only if

$$E'(N_{n1}^* - 1)^2 = 1 + o(n^{-1/2}) \quad \text{and} \quad E'(N_{n1}^* - 1)^3 = 1 + o(1) \tag{1.2}$$

as n increases. Here, E' denotes expectation conditional on the data. [The percentile method may be used to obtain second-order accurate approximations if the sampling variance is known, but when variance is unknown, percentile-t is an appropriate choice. See Hall (1992), page 92ff. In later work we shall drop both the $'$ from E', and the n and $*$ from N_{ni}^*, so as to make our notation less cumbersome.]

Of course, the exchangeability condition demands as well that

$$E'(N_{n1}^*) = 1. \tag{1.3}$$

Thus, the very elementary conditions (1.2) and (1.3) are seen as being the main requirements of bootstrap algorithms which successfully capture second-order features of the target distribution. General bootstrap algorithms which satisfy those assumptions, but are not necessarily restricted to the classical "sampling with replacement" prescription, can be expected to produce distributional approximations whose accuracy is as good as that associated with the latter scheme. Such resampling plans are discussed by Haeusler, Mason and Newton (1992) and Praestgaard and Wellner (1993). In Haeusler, Mason and Newton (1992), higher-order properties of a modification of the wild bootstrap are studied. Praestgaard and Wellner (1993) give sufficient conditions for consistency of resampling empirical processes. Hušková and Janssen (1993a, b) have developed generalized bootstrap methods for U-statistics.

2. Methods and theory.

2.1. *Summary.* Section 2 introduces a general approach to resampling methods, develops a theory to describe its properties and draws conclusions from that theory. The methodology is described in Section 2.2, where it is considered from the viewpoint of the generation of random measures. We show that in most circumstances, our general methods may be treated as a straightforward resampling scheme, although there are some instances where that is not appropriate. They involve applications of the wild bootstrap to nonlinear statistics, and are discussed in Sections 2.5 and 2.6. Sections 2.3 to 2.6 treat the special case of the mean and the Studentized mean. Asymptotic theory for other cases, such as smooth functions of means, is virtually identical.

Section 2.3 introduces notation in the case of the mean, and Section 2.4 states theorems about moments and cumulants of means and exchangeable variables. These formulae are required for analysing general resampling schemes, and are applied to that purpose in Sections 2.5 and 2.6. There we compare the asymptotic properties of different resampling methods for constructing distribution estimators, and describe the coverage accuracies of confidence intervals derived from those estimators.

In Section 2.4 we provide a simple, necessary and sufficient condition for general bootstrap distribution estimators to be accurate up to and including terms of order $n^{-1/2}$. See condition (2.2). Sections 2.5 and 2.6 develop this line of argument further, and show that under mild and simple conditions on the form of the general bootstrap approximation, the latter differs from the true distribution only in terms of size n^{-1}. Furthermore, the difference may be written as $n^{-1}\psi(x)\phi(x) + O(n^{-3/2})$, where ψ is an odd, third-degree polynomial and ϕ is the $N(0,1)$ density. Explicit formulae for ψ are developed for a variety of different versions of the general bootstrap, and lead to important qualitative conclusions about relative lengths and coverages of different intervals.

All the comparisons of distributions that we make are in terms of moments or cumulants. This approach is standard in the literature—see, for example, Chapters 3 and 6 of Kendall, Stuart and Ord (1987). In particular, it leads directly to expansions of characteristic and moment generating functions, and so to descriptions of the order of approximation in well-known limit theorems for those quantities. It also leads to accounts of the order of approximation in many other settings. Indeed, if Z_n converges to Z in distribution as $n \to \infty$, then from knowledge of the way in which moments of Z_n compare with those of Z we may deduce the order of approximation of $E\{f(Z_n)\}$ to $E\{f(Z)\}$ for a wide variety of smooth functions f. Should it be possible to rigorously verify the standard formal inversions of expansions of characteristic functions, and thereby obtain analogous expansions of distribution functions, then our results immediately produce properties about the order of approximation in that setting. However, direct comparison of distributions via cumulant expansions is very informative and relevant even in the absence of expansions of the cumulative distribution functions of those distributions.

2.2. *A general resampling scheme.* In this section we introduce a class of resampling plans for (real-valued) statistical functionals T. Consider first the usual bootstrap procedure. For brevity let us discuss here first only the case of non-Studentized statistics. Given an independent and identically distributed (i.i.d.) sample $\mathcal{X} = \{X_1, \dots, X_n\}$ with distribution P, the bootstrap provides an estimate for the distribution of $T(\widehat{P}_n) - T(P)$, where $\widehat{P}_n$ is the empirical distribution based on $\mathcal{X}$. Usually the bootstrap estimate is introduced as the conditional distribution of $T(\widehat{P}_n^*) - T(\widehat{P}_n)$, where $\widehat{P}_n^*$ is the empirical distribution of a sample drawn from $\widehat{P}_n$.

There exists another interpretation of the bootstrap procedure which will lead below naturally to a more general class of resampling plans. Given the sample $\mathcal{X}$, the measure $\widehat{P}_n^*$ is a random distribution with weights at the points $X_1, \dots, X_n$. This viewpoint can also be used for the wild bootstrap, which has hitherto been introduced for linear statistics T. Let us briefly recall the definition. Suppose for simplicity that $T(P) = \int x\,dP$. Then the wild bootstrap proceeds as follows. First one generates n independent and identically distributed random variables $R^*_{n1}, \dots, R^*_{nn}$ with $ER^*_{n1} = 0$ and $E(R^*_{n1})^2 = E(R^*_{n1})^3 = 1$. Then the distribution of $T(\widehat{P}_n) - T(P)$ is estimated by the (conditional) distribution of $\Sigma_{i=1}^n R^*_{ni}(X_i - \overline{X}_1)$, where $\overline{X}_1$ is the sample mean.

The wild bootstrap is also suited to the case of nonidentically distributed, but independent, observations $X_1, \dots, X_n$. The idea behind this procedure is that the conditional distribution of $R^*_{ni}(X_i - \overline{X}_1)$ may be interpreted as an estimate of the distribution of $X_i - EX_i$. We shall use a similar idea. However, as indicated above, we shall describe our resampling not as a drawing of independent observations, but as the generation of random measures.

We use the following "wild" heuristics. We seek an estimate of the distribution of the random measures $Q_i \equiv \delta_{X_i} - P$, where δ_x is the point mass in x. For the distribution of Q_i we propose as an estimate the conditional distribution of $\widehat{Q}_i \equiv N^*_{ni}(\delta_{X_i} - \widehat{P}_n)$, where $N^*_{n1}, \dots, N^*_{nn}$ is a sequence of random variables.

The resampling operation should reflect the stationarity and order-invariance of $\mathcal{X}$, in the sense that the variables N^*_{ni}, that is, the number of times that X_i appears in the resample, should form an exchangeable sequence. Nevertheless, we do not require the full force of the exchangeability assumption, but employ only moment versions of it; see conditions (B: m)–(D: m) in Section 2.4. The reader is referred to Taylor, Daffer and Patterson (1985) for a detailed account of exchangeability.

In analogy to $\widehat{P}_n = P + n^{-1}\Sigma_{1 \le i \le n} Q_i$, we define $\widehat{P}^*_n = \widehat{P}_n + n^{-1}\Sigma_{1 \le i \le n}\widehat{Q}_i$. This gives

$$\widehat{P}^*_n = n^{-1}\sum_{i=1}^{n}\left(N^*_{ni} + 1 - n^{-1}N^*_n\right)\delta_{X_i},$$

where $N^*_n = \Sigma_{j=1}^n N^*_{nj}$. In particular, for the special case where $N^*_n = n$, this reduces to $\widehat{P}^*_n = n^{-1}\Sigma_{1\le i \le n} N^*_{ni}\delta_{X_i}$. In general, we need not assume that the N^*_{nj}'s add up to n. Furthermore, $\widehat{P}^*_n$ need not be a positive measure. We ask only that $\int d\widehat{P}^*_n = 1$.

As our estimate of the distribution of $T(\widehat{P}_n) - T(P)$ we propose the conditional distribution of $T(\widehat{P}^*_n) - T(\widehat{P}_n)$. Furthermore, the Studentized functional $S^{-1}(\widehat{P}_n)\ \{T(\widehat{P}_n) - T(P)\}$ can be estimated by the conditional distribution of $S^{-1}(\widehat{P}^*_n)\{T(\widehat{P}^*_n) - T(\widehat{P}_n)\}$. Here, $S^2(P)$ is a functional which approximates the variance of $T(\widehat{P}_n) - T(P)$ under P.

Let us now briefly mention special choices of the distribution of $(N^*_{n1}, \dots, N^*_{nn})$.

EXAMPLE 1 (Bootstrap). Here, $(N^*_{n1}, \dots, N^*_{nn})$ has a multinomial distribution with parameters $(n^{-1}, \dots, n^{-1}; n)$.

EXAMPLE 2 (Bootstrap, revised). As proposed for instance by Bickel and Freedman (1981), let us consider the bootstrap with resample size M^*_n different from n. Suppose X_i appears just M^*_{ni} times in this resample. More explicitly, we might suppose that M_n^* is a random variable and that given $M_n^* = m$ the n-tuple $(M_{n1}^*, \dots, M_{nn}^*)$ has a multinomial distribution with parameters $(n^{-1}, \dots, n^{-1}; m)$. To recover condition (1.1) we might put $N_{ni}^* = M_{ni}^*/M_n^*$. In the case of resampling the mean, this procedure corresponds to using the conditional distribution of

$$(M_n^*)^{-1} \sum_{i=1}^{n} M_{ni}^* X_i,$$

instead of

$$n^{-1} \sum_{i=1}^{n} M_{ni}^* X_i,$$

to approximate the distribution of the sample mean.

If the statistic $T(\widehat{P}_n)$ is not a mean, then the approaches described in the next two examples may only be treated strictly as resampling methods if additional values are appended to the sample. In particular, when studying the Studentized mean using the wild bootstrap, the mean of the squares (needed to calculate the variance) is taken to be the mean of the values of $(N_{ni}^* X_i)^2$ rather than the mean of the values of $N_{ni}^* X_i^2$.

EXAMPLE 3 (Wild bootstrap). Define $N_{ni}^* \equiv R_{ni}^* + 1$, where the R_{ni}^*'s are i.i.d. with $E(R_{ni}^*) = 0$ and $E(R_{ni}^*)^2 = E(R_{ni}^*)^3 = 1$. In the case of the mean, this amounts to resampling in such a way that the numbers of appearances of sample values are i.i.d., and the resample size is $n + \Sigma R_{ni}^*$. For more general statistics, the random measure approach discussed above provides a way of defining the wild bootstrap in contexts where it has not previously been considered.

EXAMPLE 4 (Wild bootstrap, revised). We propose the following modification of the wild bootstrap. As before, define $(R_{n1}^*, \dots, R_{nn}^*)$ to be i.i.d., but put $N_{ni}^* \equiv n(1 + R_{ni}^*)/(n + R_n^*)$, where $R_n^* \equiv \Sigma_{1 \le i \le n} R_{ni}^*$. For positive random variables R_{ni}^*, this resampling procedure ensures that P_n^* is a proper probability measure.

EXAMPLE 5 (Generalized jackknife). A resampling plan which fits in our framework and which uses a minimal "amount of randomness" can be described as follows. Choose a (deterministic) vector $\rho = (\rho(1), \dots, \rho(n))$ in $\mathbb{R}^n$. Then put $N_{ni}^* \equiv \rho(\Pi(i))$, where Π is a random permutation of $(1, \dots, n)$.

2.3. *Resampling linear statistics.* To simplify matters in this section and in the remainder of the paper, we consider only the mean $T(P) = \int x \, dP$. Asymptotic theory for other cases is virtually identical, in qualitative terms, and so there is little point in complicating matters by treating the general case explicitly. The scale functional S is always chosen as the variance functional $S(P)^2 = \int \{x - T(P)\}^2 \, dP$.

With $\overline{X}_j \equiv n^{-1}\Sigma_{1\le i\le n}X_i^j$ and $\overline{X}_j^* \equiv n^{-1}\Sigma_{1\le i\le n}N_{ni}^*X_i^j$, we have now

$$U_j^* \equiv \int x^j(d\widehat{P}_n^* - d\widehat{P}_n) = n^{-1}\sum_{i=1}^{n}(N_{ni}^* - n^{-1}N_n^*)X_i^j$$
$$= n^{-1}\sum_{i=1}^{n}(N_{ni}^* - 1)(X_i^j - \overline{X}_j).$$

The sample and resample variance are given by

$$\begin{aligned}\widehat{\sigma}^2 &\equiv S^2(\widehat{P}_n) = \overline{X}_2 - \overline{X}_1^2,\\ \widehat{\sigma}^{*2} &\equiv S^2(\widehat{P}_n^*) = n^{-1}\sum_{i=1}^{n}(N_{ni}^* + 1 - n^{-1}N_n^*)(X_i - \overline{X}_1)^2\\ &\quad - \left\{n^{-1}\sum_{i=1}^{n}(N_{ni}^* + 1 - n^{-1}N_n^*)(X_i - \overline{X}_1)\right\}^2\\ &= \widehat{\sigma}^2 + U_2^* - 2\overline{X}_1U_1^* - U_1^{*2},\end{aligned}$$

respectively. Let μ and σ^2 denote the population mean and variance. Then the resample estimate of the distribution of $U \equiv \overline{X}_1 - \mu$ is given by the conditional distribution of $U^* \equiv \overline{X}_1^* - \overline{X}_1$, and the resample estimate of the distribution of $V \equiv \widehat{\sigma}^{-1}(\overline{X}_1 - \mu)$ is given by the conditional distribution of $V^* \equiv \widehat{\sigma}^{*-1}(\overline{X}_1^* - \overline{X}_1)$.

2.4. *Moments and cumulants of weighted sums of exchangeable variables.* Recall the definitions of U, V, U^* and V^* given in Section 2.3. The asymptotic properties of the distributions of U and U^*, and of V and V^*, may be compared in terms of their Edgeworth expansions, or, usually equivalently, via their cumulant expansions. In this section we shall develop approximate formulae for those quantities, and also for their counterparts where the analogues of the N_{ni}^*'s are independent of one another but with the same marginal distribution as the exchangeable sequence $\{N_{ni}^*\}$. It turns out to be convenient to describe formulae for cumulants by comparing the exchangeable and independent cases.

In this analysis we are conditioning on the sequence $\{X_i\}$, and so we regard those variables as constants. They are no more than weights for the general exchangeable sequence $\{N_{ni}^*,\ 1 \le i \le n\}$, and indeed we may allow the weights to form a triangular array, $\{v_{ni},\ 1 \le i \le n < \infty\}$, rather than a linear array $\{X_i,\ 1 \le i < \infty\}$. (When returning to the special cases of U^* and V^* we take $v_{ni} \equiv X_i$.) For convenience we drop the asterisk from N_{ni}^*. Recall that we assume

$E(N_{ni}) = 1$. We ask that for an integer $m \geq 1$, and $n \geq r + j$,

$$\text{(A: } m\text{)} \qquad \sup_{n\geq 1} n^{-1} \sum_{i=1}^{n} |v_{ni}|^m < \infty,$$

$$\text{(B: } m\text{)} \qquad \sup_{n\geq 1} E|N_{n1}|^m < \infty,$$

$$\text{(C: } m\text{)} \qquad \begin{aligned} &E\{(N_{n1}-1)^{s_1}\cdots(N_{nr}-1)^{s_r}(N_{n,r+1}-1)\cdots(N_{n,r+j}-1)\} \\ &\quad = O\{n^{-1}I(j=1,2)+n^{-2}I(j\geq 3)\} \end{aligned}$$

for all integers $r \geq 0, 2 \leq s_1 \leq \cdots \leq s_r$ and $j \geq 1$ satisfying $s_1 + \cdots + s_r + j = m$, and

$$\text{(D: } m\text{)} \qquad E\{(N_{n1}-1)^{s_1}\cdots(N_{nr}-1)^{s_r}\} - \prod_{i=1}^{r} E(N_{ni}-1)^{s_i} = O(n^{-1})$$

for all integers $r \geq 1$ and $2 \leq s_1 \leq \cdots \leq s_r$ satisfying $s_1 + \cdots + s_r = m$. In these formulae, expectation is taken in the distribution of the N_{ni}'s, with the v_{ni}'s regarded as fixed. Translating to the bootstrap context, this is equivalent to taking expectation conditional on the data $\mathcal{X}$.

When $v_{ni} = X_i$, condition (A: m) holds with probability 1 (in the distribution of $X_1, X_2, \ldots$) provided that $E|X|^m < \infty$.

For the sake of simplicity, drop the subscript n from N_{ni} and v_{ni}, and define $\bar{v}_j \equiv n^{-1}\Sigma v_k^j$, $w_i \equiv v_i - \bar{v}_1$, $\bar{w}_j \equiv n^{-1}\Sigma w_i^j$, $t^2 \equiv \bar{v}_2 - \bar{v}_1^2$, $\nu_j \equiv E(N_1 - 1)^j$,

$$\begin{aligned}
\alpha_{r1} &\equiv n\Big[E\{(N_1-1)^2\cdots(N_r-1)^2\} - \nu_2^r\Big], \\
\alpha_{r-1,2} &\equiv nE\{(N_1-1)^2\cdots(N_{r-2}-1)^2(N_{r-1}-1)(N_r-1)\}, \\
S_1 &\equiv n^{-1}\sum_{i=1}^{n}(N_i-1)w_i, \qquad S_2 \equiv n^{-1}\sum_{i=1}^{n}(N_i-1)(v_i^2-\bar{v}_2), \\
\tau^2 &\equiv t^2 + S_2 - 2\bar{v}_1 S_1 - S_1^2, \\
T &\equiv t\tau^{-1}S_1, \\
T_1 &\equiv S_1\{1 - \tfrac{1}{2}t^{-2}(S_2 - 2\bar{v}_1S_1) + \tfrac{1}{2}t^{-2}S_1^2 + \tfrac{3}{8}t^{-4}(S_2-2\bar{v}_1S_1)^2\}.
\end{aligned}$$

In the event that $\Sigma N_i = n$ we have an equivalent definition of τ:

$$\tau^2 = n^{-1}\sum_{i=1}^{n} N_i v_i^2 - \left(n^{-1}\sum_{i=1}^{n} N_i v_i\right)^2.$$

Then S_1, T represent $U_1^*, \hat{\sigma}V^*$, respectively. We shall develop formulae for moments and cumulants of $n^{1/2}S_1$ and $n^{1/2}T$ up to terms of order $n^{-3/2}$, and also for the case where the N_i's are independent. To describe the latter context, let $N_{01}, \ldots, N_{0n}$ denote independent and identically distributed random variables with the distribution of N_1, and put

$$S_{01} \equiv n^{-1}\sum_{i=1}^{n}(N_{0i}-1)w_i, \qquad S_{02} \equiv n^{-1}\sum_{i=1}^{n}(N_{0i}-1)\left(v_i^2-\bar{v}_2\right),$$

$$T_{01} \equiv S_{01}\left\{1-\tfrac{1}{2}t^{-2}(S_{02}-2\bar{v}_1S_{01})+\tfrac{1}{2}t^{-2}S_{01}^2+\tfrac{3}{8}t^{-4}(S_{02}-2\bar{v}_1S_{01})^2\right\}.$$

Our main result follows. All proofs of theorems are deferred to Section 3.

THEOREM 2.1. *Let $k \geq 1$ be an integer.*

(i)(a) *If conditions* (A: k)–(D: k) *hold, then*

$$E\left\{\left(n^{1/2}S_1\right)^k\right\}-E\left\{\left(n^{1/2}S_{01}\right)^k\right\}$$
$$=\begin{cases} n^{-1}\overline{w}_2^{\,l}\dfrac{(2l)!}{l!2^l}(\alpha_{l,1}-l\alpha_{l,2})+O\left(n^{-2}\right), & \text{for } k=2l,\\ O\left(n^{-3/2}\right), & \text{for } k=2l+1.\end{cases}$$

(b) *If conditions* (A: $3k$)–(D: $3k$) *hold, then*

$$n^{1/2}T = n^{1/2}T_1 + O_p\left(n^{-3/2}\right) \tag{2.1}$$

and

$$E\left\{\left(n^{1/2}T_1\right)^k\right\}-E\left\{\left(n^{1/2}T_{01}\right)^k\right\}$$
$$=E\left\{\left(n^{1/2}S_1\right)^k\right\}-E\left\{\left(n^{1/2}S_{01}\right)^k\right\}+\begin{cases} O\left(n^{-2}\right), & \text{for } k=2l,\\ O\left(n^{-3/2}\right), & \text{for } k=2l+1.\end{cases}$$

(ii) *Let* (C′: m) *denote the version of* (C: m) *in which the right-hand side is replaced by* $o\{n^{-1/2}I(j=1,2)+n^{-3/2}I(j\geq 3)\}$, *and* (D′: m) *the version of* (D: m) *in which the right-hand side is replaced by* $o(n^{-1/2})$.

(a) *If* (A: k), (B: k), (C′: k) *and* (D′: k) *hold, then*

$$E\left\{\left(n^{1/2}S_1\right)^k\right\}-E\left\{\left(n^{1/2}S_{01}\right)^k\right\}=o\left(n^{-1/2}\right).$$

(b) *If* (A: $3k$), (B: $3k$), (C′: $3k$) *and* (D′: $3k$) *hold, then* (2.1) *is true, and*

$$E\left\{\left(n^{1/2}T_1\right)^k\right\}-E\left\{\left(n^{1/2}T_{01}\right)^k\right\}=o\left(n^{-1/2}\right).$$

The moments and cumulants of S_{01} and T_{01} are derivable by direct calculation, as we now outline. Let $\kappa_j(Z)$ denote the jth cumulant of a random variable Z. Since S_{01} and T_{01} may be represented as smooth functions of means of independent sequences, then traditional arguments may be used to show that for

$Z = n^{1/2}S_{01}$ or $Z = n^{1/2}T_{01}$,

$$\kappa_k(Z) = \begin{cases} O(n^{-3/2}), & \text{odd } k \geq 5, \\ O(n^{-2}), & \text{even } k \geq 6. \end{cases}$$

See, for example, Hall (1992), Section 2.4, especially Theorem 2.1. Since we are interested only in departures from normality up to and including terms of order n^{-1}, then it suffices to develop formulae for the cumulants of $n^{1/2}S_{01}$ and $n^{1/2}T_{01}$ up to that order. These formulae are given below.

THEOREM 2.2. *Under condition* (A: 4),

$$\kappa_1(n^{1/2}S_{01}) = 0, \qquad \kappa_2(n^{1/2}, S_{01}) = \nu_2 t^2, \qquad \kappa_3(n^{1/2}S_{01}) = n^{-1/2}\nu_3\overline{w}_3,$$
$$\kappa_4(n^{1/2}S_{01}) = n^{-1}(\nu_4 - 3\nu_2^2)\overline{w}_4 + O(n^{-2}).$$

Under condition (A: 12),

$$\begin{aligned}
\kappa_1(n^{1/2}T_{01}) &= -n^{-1/2}\tfrac{1}{2}\nu_2 t^{-2}\overline{w}_3 + O(n^{-3/2}),\\
\kappa_2(n^{1/2}T_{01}) &= \nu_2 t^2 + n^{-1}\Big\{\nu_3(t^2 - t^{-2}\overline{w}_4) + \nu_2^2(2t^2 + \tfrac{7}{4}t^{-4}\overline{w}_3^2 + t^{-2}\overline{w}_4)\Big\}\\
&\quad + O(n^{-2}),\\
\kappa_3(n^{1/2}T_{01}) &= n^{-1/2}(\nu_3 - 3\nu_2^2)\overline{w}_3 + O(n^{-3/2}),\\
\kappa_4(n^{1/2}T_{01}) &= n^{-1}\Big\{(\nu_4 - 3\nu_2^2)\overline{w}_4 + 6\nu_2\nu_3(t^4 - t^{-2}\overline{w}_3^2 - \overline{w}_4)\\
&\qquad + 3\nu_2^3(3t^4 + 6t^{-2}\overline{w}_3^2 + \overline{w}_4)\Big\} + O(n^{-2}).
\end{aligned}$$

Next we comment on the extent to which the cumulant expansions of $n^{1/2}S_1/t$ agree with those of $n^{1/2}U/\sigma = n^{1/2}(\overline{X}_1 - \mu)/\sigma$. By the first part of Theorem 2.2,

$$\kappa_1(n^{1/2}S_{01}/t) = 0, \qquad \kappa_2(n^{1/2}S_{01}/t) = \nu_2, \qquad \kappa_3(n^{1/2}S_{01}/t) = n^{-1/2}\nu_3 t^{-3}\overline{w}_3,$$

and by Theorem 2.1, these expansions agree with those of the cumulants of $n^{1/2}S_1/t$ up to terms of smaller order than $n^{-1/2}$. By direct calculation,

$$\kappa_1(n^{1/2}U/\sigma) = 0, \qquad \kappa_2(n^{1/2}U/\sigma) = 1, \qquad \kappa_3(n^{1/2}U/\sigma) = n^{-1/2}\gamma,$$

where $\gamma = E(X - \mu)^3$. Comparing the last two lines of displayed formulae, and noting that

$$t^{-3}\overline{w}_3 = \hat{\sigma}^{-3}n^{-1}\sum_{i=1}^{n}(X_i - \overline{X}_1)^2 = \gamma + o(1),$$

we see that the first three cumulants of $n^{1/2}U/\sigma$ and $n^{1/2}S_1/t$ agree to $o(n^{-1/2})$ if and only if

$$\nu_2 = 1 + o(n^{-1/2}), \qquad \nu_3 = 1 + o(1). \tag{2.2}$$

Since fourth- and higher-order cumulants of $n^{1/2}U/\sigma$ and $n^{1/2}S_1/t$ are of order n^{-1}, then (2.2) is necessary and sufficient for agreement of all cumulants to $o(n^{-1/2})$. The agreement cannot be to $O(n^{-1})$, since (for example) the fact that $t^{-3}\overline{w}_3$ and γ are a distance $n^{-1/2}$ apart means that third cumulants cannot agree to $O(n^{-1})$.

Very similar comments apply to a comparison of the distributions of $n^{1/2}T/t$ and $n^{1/2}V = n^{1/2}(\overline{X}_1 - \mu)/\widehat{\sigma}$. The cumulants of the asymptotic distribution of $n^{1/2}T/t$ agree with those of $n^{1/2}T_{01}/t$ up to terms of $o(n^{-1/2})$, and by Theorem 2.2, the latter cumulants are given by

$$\begin{aligned}
\kappa_1\left(n^{1/2}T_{01}/t\right) &= -n^{-1/2}\tfrac{1}{2}\nu_2 t^{-3}\overline{w}_3 + o\left(n^{-1/2}\right),\\
\kappa_2\left(n^{1/2}T_{01}/t\right) &= \nu_2 + o\left(n^{-1/2}\right),\\
\kappa_3\left(n^{1/2}T_{01}/t\right) &= n^{-1/2}\left(\nu_3 - 3\nu_2^2\right)t^{-3}\overline{w}_3 + o\left(n^{-1/2}\right);
\end{aligned}$$

and by direct calculation, the first three cumulants of the asymptotic distribution of $n^{1/2}V$ are $-n^{-1/2}\frac{1}{2}\gamma + o(n^{-1/2})$, $1 + o(n^{-1/2})$, $-n^{-1/2}2\gamma + o(n^{-1/2})$, respectively. Therefore, (2.2) is again necessary and sufficient for agreement of these cumulants to $o(n^{-1/2})$. And since fourth- and higher-order asymptotic cumulants are of size $o(n^{-1/2})$, this conclusion extends from agreement of the first three cumulants to the agreement of cumulants of all orders. Once again, agreement to $O(n^{-1})$, rather than simply $o(n^{-1/2})$, is not possible.

The argument above is quite general, in that it does not require any assumptions about the numerical values of the quantities α_{r1} and α_{r2}, introduced just prior to Theorem 2.1. However, if we wish to examine terms of size n^{-1} in cumulant expansions, then the α's play a crucial role. They will almost always contribute terms of size n^{-1} to the fourth cumulant, and will also, in many instances, contribute terms of that size to sixth and higher even-indexed cumulants. Cases where this does not happen include that where the N_{ni}^*'s are independent (the so-called wild bootstrap; see Section 2.2), and that where the N_{ni}^*'s have a distribution which, conditional on ΣN_{ni}^*, is a scale multiple of a multinomial with each probability equal to n^{-1}. (This produces the usual bootstrap with general sample size; again, see Section 2.2.) In these instances it may be shown [e.g., Hall (1992), Section 2.4, especially Theorem 2.1] that even cumulants higher than the fourth are of order n^{-2}.

2.5. *Comparison of different resampling methods.* To simplify our comparison, let us assume that we are in a situation such as that described just above, where the only terms of size n^{-1} that enter into formulae for asymptotic cumulants of $n^{1/2}S_1$ and $n^{1/2}T$ come from second and fourth cumulants. Thus, all contributions to the distributions of $n^{1/2}S_1$ and $n^{1/2}T$, deriving from fifth and higher cumulants, are of size $o(n^{-1})$.

It is particularly convenient to consider all formulae relative to their counterparts in the case of the common bootstrap with sample size n. There, $(N_{n1}^*, \dots, N_{nn}^*)$ are multinomial with probabilities $(n^{-1}, \dots, n^{-1})$, and $\Sigma N_{ni}^* = n$. In this case it may be shown that

$$\nu_2 = 1 - n^{-1}, \qquad \nu_3 = 1 + O(n^{-1}), \qquad \nu_4 = 4 + O(n^{-1}),$$
$$\alpha_{12} = -1 + O(n^{-1}), \qquad \alpha_{21} = -1 + O(n^{-1}), \qquad \alpha_{22} = -1 + O(n^{-1}).$$

For more general bootstrap methods, such as those considered in the examples of Section 2.2, formulae differ from those above only in terms of order n^{-1} in the case of ν_j, and $o(1)$ in the case of α_{ij}. Thus, the analogues of formulae in the previous display are

$$\begin{gathered} \nu_2 = 1 + (\beta_1 - 1)n^{-1} + o(n^{-1}), \\ \nu_3 = 1 + o(n^{-1/2}), \qquad \nu_4 = \beta_2 + 4 + o(1), \\ \alpha_{11} = 0, \qquad \alpha_{12} = \beta_3 - 1, \qquad \alpha_{21} = \beta_4 - 1 + o(1), \qquad \alpha_{22} = \beta_5 - 1 + o(1), \end{gathered} \tag{2.3}$$

where the constants $\beta_1, \dots, \beta_5$ are all 0 in the case of the common bootstrap. The identities (2.3) amount to definitions of $\beta_1, \dots, \beta_5$. Note that $\alpha_{11} = 0$ in all circumstances, and that calculation of cumulant expansions to order n^{-1} requires ν_3 only to order $n^{-1/2}$ and ν_4 to order 1.

When discussing bootstrap methods using the results in Section 2.4, we should, as noted earlier, take $v_i = X_i$ and $w_i = X_i - \overline{X}_1$, and use expectations conditional on the X_i's. That notation is employed below. Let E^B and κ_j^B denote expectation and jth cumulant, respectively, in the case of the common bootstrap. Theorems 2.1 and 2.2, and the standard formulae for cumulants in terms of moments, may be used to show that, with $\widehat{\sigma}^2 = n^{-1}\Sigma(X_i - \overline{X}_1)^2$ and $\widehat{\mu}_4 = n^{-1}\Sigma(X_i - \overline{X}_1)^4$,

$$\begin{aligned}
\kappa_1(n^{1/2}S_1) - \kappa_1^B(n^{1/2}S_1) &= E(n^{1/2}S_1) - E^B(n^{1/2}S_1) = o(n^{-1}), \\
\kappa_2(n^{1/2}S_1) - \kappa_2^B(n^{1/2}S_1) &= E(n^{1/2}S_1)^2 - (En^{1/2}S_1)^2 \\
&\quad - \Big\{E^B(n^{1/2}S_1)^2 - (E^B n^{1/2}S_1)^2\Big\} \\
&= n^{-1}\widehat{\sigma}^2(\beta_1 - \beta_3) + o(n^{-1}), \\
\kappa_3(n^{1/2}S_1) - \kappa_3^B(n^{1/2}S_1) &= E(n^{1/2}S_1)^3 - 3E(n^{1/2}S_1)E(n^{1/2}S_1)^2 + 2(En^{1/2}S_1)^3 \\
&\quad - \Big\{E^B(n^{1/2}S_1)^3 - 3(E^B n^{1/2}S_1)E^B(n^{1/2}S_1)^2 \\
&\qquad + 2(E^B n^{1/2}S_1)^3\Big\} \\
&= o(n^{-1}), \\
\kappa_4(n^{1/2}S_1) - \kappa_4^B(n^{1/2}S_1) &= E(n^{1/2}S_1 - En^{1/2}S_1)^4 - 3\Big\{\kappa_2(n^{1/2}S_1)\Big\}^2 \\
&\quad - \Big[E^B(n^{1/2}S_1 - En^{1/2}S_1)^4 - 3\Big\{\kappa_2^B(n^{1/2}S_1)\Big\}^2\Big] \\
&= n^{-1}\big\{\widehat{\sigma}^4 3(\beta_4 - 2\beta_5) - \widehat{\sigma}^4 6(\beta_1 - \beta_3) + \widehat{\mu}_4\beta_2\big\} + O(n^{-2}) \\
&= n^{-1}\widehat{\sigma}^4\big\{3(2\beta_3 + \beta_4 - 2\beta_1 - 2\beta_5) + \widehat{\sigma}^{-4}\widehat{\mu}_4\beta_2\big\} \\
&\quad + O(n^{-2}).
\end{aligned}$$

[The remainder terms here and below are of the stated orders uniformly in samples for which (A: 4) holds for an arbitrary but fixed bound on the right-hand side. As explained shortly after (A: 4), this is true "with probability 1, for all sufficiently large n," if an appropriate moment condition is valid.] Similarly,

$$\begin{aligned}
\kappa_1(n^{1/2}T_1) - \kappa_1^B(n^{1/2}T_1) &= o(n^{-1}),\\
\kappa_2(n^{1/2}T_1) - \kappa_2^B(n^{1/2}T_1) &= n^{-1}\hat\sigma^2(\beta_1-\beta_3) + o(n^{-1}),\\
\kappa_3(n^{1/2}T_1) - \kappa_3^B(n^{1/2}T_1) &= o(n^{-1}),\\
\kappa_4(n^{1/2}T_1) - \kappa_4^B(n^{1/2}T_1) &= n^{-1}\hat\sigma^4\{3(2\beta_3+\beta_4-2\beta_1-2\beta_5)\hat\sigma^{-4}\hat\mu_4\beta_2\}\\
&\quad + O(n^{-2}).
\end{aligned}$$

Expanding the characteristic functions of $n^{1/2}S_1/\hat\sigma$ and $n^{1/2}T/\hat\sigma$ and supposing that those formulae may be inverted in the obvious manner to produce expansions of distribution functions, we deduce that the difference between the (conditional) distribution functions of $n^{1/2}S_1/\hat\sigma$ (or $n^{1/2}T/\hat\sigma$) under a regime satisfying (2.3), and the common bootstrap, is given by $\phi(x)$ [the $N(0,1)$ density] multiplied by

$$\begin{aligned}
&-n^{-1}x\left[\tfrac{1}{2}(\beta_1-\beta_3)+\tfrac{1}{24}(x^2-3)\{3(2\beta_3+\beta_4-2\beta_1-2\beta_5)+\hat\sigma^{-4}\hat\mu_4\beta_2\}\right]\\
&\quad = n^{-1}\tfrac{1}{8}x\Big\{\hat\sigma^{-4}\hat\mu_4\beta_2+10\beta_3+3\beta_4-4\beta_1-6\beta_5\\
&\qquad\qquad + x^2\big(2\beta_1-\tfrac{1}{3}\hat\sigma^{-4}\hat\mu_4\beta_2-2\beta_3-\beta_4+2\beta_5\big)\Big\}\\
&\quad = n^{-1}\psi(x),
\end{aligned}\tag{2.4}$$

say. Note particularly that the quantities $\hat\sigma^2$ and $\hat\mu_4$ appearing in the definition of ψ depend on the observed data; they were defined earlier in this paragraph. Furthermore, the β's may depend on the data.

Perhaps the simplest form of the wild bootstrap is that where the N_i's are independent with $E(N_i)=1$ and $E(N_i-1)^2 = E(N_i-1)^3 = 1$. Here, (2.3) holds with $\beta_1=\beta_3=\beta_4=\beta_5=1$, and β_2 is undetermined. The function ψ now has the form

$$\psi(x) = -\tfrac{1}{24}x(x^2-3)(\hat\sigma^{-4}\hat\mu_4\beta_2-3).\tag{2.5}$$

In particular, if $\beta_2 < 3\hat\sigma^4\hat\mu_4^{-1}$, the wild bootstrap has lighter tails than the common bootstrap. Note that $\beta_2 = 0$ if the fourth cumulant of N_1 equals unity.

We should remark that the version of the wild bootstrap considered here differs slightly from that considered by Liu (1988) and Mammen (1992), in that a different variance estimate is employed. Our estimate is given by

$$\tau^2 = t^2 + n^{-1}\sum_{i=1}^{n} N_i(w_i^2-t^2) - S_1^2,$$

whereas that suggested by Liu and Mammen is

$$\rho^2 = n^{-1}\sum_{i=1}^{n}(N_i-1)^2 w_i^2 - S_1^2.$$

The resulting Studentized statistic is $R = t\rho^{-1}S_1$, which is the analogue of T. Its Taylor approximant, the analogue of T_1, is

$$R_1 = S_1\left(1 - \tfrac{1}{2}t^{-2}S_2 + \tfrac{1}{2}t^{-2}S_1^2 + \tfrac{3}{8}t^{-4}S_3^2\right) = R + O_p\left(n^{-2}\right),$$

where $S_3 = n^{-1}\Sigma_{1\le i\le n}\{(N_i - 1)^2 - 1\}w_i^2$. The cumulants of R_1 may be derived as were those of T_1. In this way it may be shown that the first three cumulants of $n^{1/2}R_1$ are identical to their counterparts for T_1 in the case of the common bootstrap:

$$\kappa_j\left(n^{1/2}R_1\right) - \kappa_j^B\left(n^{1/2}T_1\right) = \begin{cases} O\left(n^{-3/2}\right), & \text{for } j = 1, 3, \\ O\left(n^{-2}\right), & \text{for } j = 2, \end{cases}$$

and, in addition,

$$\kappa_4\left(n^{1/2}R_1\right) - \kappa_4^B\left(n^{1/2}T_1\right) = -3n^{-1}\left\{\overline{w}_4(\beta_2 + 2) + 2t^{-2}\overline{w}_3^2 + 2t^4\right\} + O\left(n^{-2}\right).$$

Thus, the distribution of R has skewness similar to that of T under the common bootstrap, but lighter tails (for all choices of β_2). The tails of R are also lighter than those of T under the wild bootstrap. This is to be expected, since the definition of ρ^2 implies that it has heavier tails than τ^2, and it appears in the denominator in the definition of R.

The "revised" form of the wild bootstrap, introduced in Example 3, may be treated in like manner. For example, if the N_i's are independent with $E(N_i) = 1$, $E(N_i - 1)^2 = E(N_i - 1)^3 = 1$ and $E(N_i^4 - 1) = \beta_2 + 4 + o(1)$, then

$$\psi(x) = -\tfrac{1}{24}x\left(x^2 - 3\right)\left(\widehat{\sigma}^{-4}\widehat{\mu}_4\beta_2 + 3\right);$$

compare (2.5). Since $x^2 - 3 < 0$ if x is an $N(0, 1)$ quantile between the levels 5% and 95%, then it follows that one-sided 95% and two-sided 90% "revised" wild bootstrap confidence intervals tend to be shorter than their counterparts for the ordinary wild bootstrap. However, since $x^2 - 3 > 0$ for $2\frac{1}{2}$% and $97\frac{1}{2}$%, or more extreme, $N(0, 1)$ quantiles, then one-sided 99% and two-sided 95% revised wild bootstrap intervals tend to be longer.

Finally, we consider the generalized jackknife, introduced in Example 4 of Section 2.2. In the notation of that example, put $\overline{\rho}_j = n^{-1}\Sigma\rho(i)^j$ and $\overline{\eta}_j = n^{-1}\Sigma\{\rho(i) - \overline{\rho}_1\}^j$. Then $E(N_i) = \overline{\rho}_1$, $\nu_2 = \overline{\rho}_2 - \overline{\rho}_1^2 = \overline{\eta}_2$, $\nu_3 = \overline{\eta}_3$. Assume without loss of generality that $\rho(i) = \rho_n(i)$ is chosen so that $\Sigma\rho(i) = n\overline{\rho}_1 = n$ (i.e., $\Sigma N_i^* = n$) and, as $n \to \infty$, $\nu_2 = 1 + (\beta_1 - 1)n^{-1} + o(n^{-1})$, $\nu_4 = \beta_2 + 4 + o(1)$, where β_1 and β_2 are constants. Then, in the notation of (2.3), $\beta_3 = \beta_5 = 0$, $\beta_4 = -(\beta_2 + 2)$. It follows that

$$\psi(x) = x\left\{\tfrac{1}{24}\left(x^2 - 3\right)\left(2\beta_1 + \beta_2 + 2 - \widehat{\sigma}^{-4}\widehat{\mu}_4\beta_2\right) - \tfrac{1}{2}\beta_1\right\}.$$

2.6. *Effect on coverage accuracy.* Let $I_1 = (-\infty, A]$ and $I_2 = [A_1, A_2]$ denote, respectively, one- and two-sided confidence intervals for the value of the unknown population mean μ. Suppose both intervals have nominal coverage probability α, and write p_i for the true coverage of I_i when a resampling scheme satisfying (2.3) is used. Let p_i^B denote the version of p_i when the common bootstrap is employed. Let ψ_0 denote the version of ψ, defined at (2.5), when $\widehat{\sigma}^{-4}\widehat{\mu}_4$ in the latter is replaced by $(\mathrm{Var}\,X)^{-2}E(X - EX)^4$. Write z_α for the α-level critical point of the $N(0,1)$ distribution; that is, z_α solves the equation $\Phi(z_\alpha) = \alpha$, where Φ is the $N(0,1)$ distribution function. We claim that

$$p_1 - p_1^B = -n^{-1}\psi_0(z_\alpha)\phi(z_\alpha) + O(n^{-3/2}), \tag{2.6}$$

$$p_2 - p_2^B = -2n^{-1}\psi_0\big(z_{(\alpha+1)/2}\big)\phi\big(z_{(\alpha+1)/2}\big) + O(n^{-2}), \tag{2.7}$$

where $\phi = \Phi'$. Thus, the difference in coverage error of both one- and two-sided confidence intervals is of size n^{-1}.

To derive (2.6) and (2.7), let $\widehat{v}_\alpha$ be the α-level quantile of the bootstrap distribution of T, and let $\widehat{v}_\alpha^B$ denote the version of $\widehat{v}_\alpha$ for the common bootstrap. Observe that, by Cornish–Fisher expansion, $\widehat{v}_\alpha - \widehat{v}_\alpha^B = -n^{-1}\psi(z_\alpha) + O_p(n^{-3/2})$. Results (2.6) and (2.7) follow from this result via the arguments of Hall (1992), Section 3.5.

The conclusions drawn in Section 2.5 about the sign of ψ, and about the relative lengths of intervals, have obvious and immediate conclusions for coverage accuracy. For example, if the common bootstrap is modified by changing resample size to $n + l$ where l is fixed, then the coverage of two-sided 95% confidence intervals tends to decrease if $l < 0$ but increase if $l > 0$.

3. Proofs of Theorems 2.1 and 2.2.

PROOF OF THEOREM 2.1. We derive only the first part of the theorem, since the second may be proved similarly. We begin by developing expansions of the moments of $n^{1/2}S_1$.

Observe that

$$E\Big\{\big(n^{1/2}S_1\big)^k\Big\} = n^{-k/2}\sum_{i_1}\cdots\sum_{i_k} w_{i_1}\cdots w_{i_k} E\Big\{\big(N_{i_1} - 1\big)\cdots\big(N_{i_k} - 1\big)\Big\}.$$

To appreciate the form of the k-fold series on the right-hand side, write $\Sigma_{i_1}\cdots\Sigma'_{i_k}$ to indicate summation over integers $i_1, \dots, i_r$ such that $1 \le i_j \le n$, for $1 \le j \le n$, and no two of $i_1, \dots, i_r$ are equal. Given $l \ge 2r$, define $\Sigma_{s_1}\cdots\Sigma_{s_r}^{(l)}$ to mean summation over integers $r \ge 1$ and $s_1, \dots, s_r$ such that $2 \le s_1 \le \cdots \le s_r$ and $s_1 + \cdots + s_r = l$. Put

$$W(s_1, \dots, s_r; j) = \sum_{i_j}\cdots\sum_{i_{r+j}}{}' w_{i_1}^{s_1}\cdots w_{i_r}^{s_r} w_{i_{r+1}}\cdots w_{i_{r+j}},$$

$$\mu(s_1, \dots, s_r; j) = E\big\{(N_1 - 1)^{s_1}\cdots(N_r - 1)^{s_r}(N_{r+1} - 1)\cdots(N_{r+j} - 1)\big\}.$$

Let $\langle x\rangle$ denote the largest integer not exceeding x. In this notation,

$$E\Big\{\big(n^{1/2}S_1\big)^k\Big\} = n^{-k/2}\sum_{j=0}^{k}\sum_{r=0}^{\langle k/2\rangle}\sum_{s_1}\cdots\sum_{s_r}^{(k-j)} D(s_1,\dots,s_r;j) \times W(s_1,\dots,s_r;j)\mu(s_1,\dots,s_r;j), \tag{3.1}$$

where the quantities $D(s_1,\dots,s_r;j)$ are combinatorial constants.

Since $\Sigma w_i = 0$ then for each $0 \le j \le k$, $W(s_1,\dots,s_r;j) = O(n^{r+j-\langle\frac{1}{2}(j+1)\rangle})$. Therefore, if $s_1+\cdots+s_r = k-j$, and each $s_i \ge 2$,

$$\begin{aligned} n^{-k/2}W(s_1,\dots,s_r;j) &= O\big(n^{-(s_1+\cdots+s_r)/2+r+j/2-\langle (j+1)/2\rangle}\big)\\ &= \begin{cases} O(1), & \text{if } j \text{ is even},\\ O\big(n^{-1/2}\big), & \text{if } j \text{ is odd}. \end{cases}\end{aligned}$$

By hypothesis, if $s_1+\cdots+s_r = k-j$, then for $j\ge 1$, $\mu(s_1,\dots,s_r;j) = O\{n^{-1}I(j=1,2) + O(n^{-2})I(j\ge 3)\}$. Hence, for $j=1$ or $j\ge 3, n^{-k/2}W(s_1,\dots,s_r;j)\mu(s_1,\dots,s_r;j) = O(n^{-3/2})$. If $j=2$ and one or more of the s_i's exceeds 2, then

$$\begin{aligned} n^{-k/2}W(s_1,\dots,s_r;j)\mu(s_1,\dots,s_r;j) &= O\big(n^{r-(s_1+\cdots+s_r)/2+j/2-\langle (j+1)/2\rangle-1}\big)\\ &= O\big(n^{-3/2}\big). \end{aligned}$$

Therefore, by (3.1),

$$\begin{aligned} E\Big\{\big(n^{1/2}S_1\big)^k\Big\} &= n^{-k/2}\sum_{r=0}^{\langle k/2\rangle}\sum_{s_1}\cdots\sum_{s_r}^{(k)} D(s_1,\dots,s_r;0)W(s_1,\dots,s_r;0)\\ &\quad\times \mu(s_1,\dots,s_r;0)\\ &\quad + I(k \text{ even})n^{-k/2}D(2,\dots,2;2)W(2,\dots,2;2)\\ &\quad\times \mu(2,\dots,2;2) + o\big(n^{-3/2}\big), \end{aligned} \tag{3.2}$$

where $r = k/2-1$ in the terms with argument $(2,\dots,2;2)$.

If one of the s_r's exceeds 2, then $n^{-k/2}W(s_1,\dots,s_r;0) = O(n^{-1/2})$. Furthermore, by hypothesis, $\mu(s_1,\dots,s_r;0) - \Pi_{1\le i\le r}\nu_{s_i} = O(n^{-1})$. It follows that if one of the s_r's exceeds 2, then

$$n^{-k/2}W(s_1,\dots,s_r;0)\mu(s_1,\dots,s_r;0) - n^{-k/2}W(s_1,\dots,s_r;0)\prod_{i=1}^{r}\nu_{s_i} = O\big(n^{-3/2}\big).$$

Hence, by (3.2),

$$
\begin{aligned}
&E\Big\{\big(n^{1/2}S_1\big)^k\Big\} - E\Big\{\big(n^{1/2}S_{01}\big)^k\Big\}\\
&\quad = I(k \text{ even})n^{-k/2}D\Big(2,2,\dots\big[\tfrac12 k \text{ times}\big];0\Big)W\Big(2,2,\dots\big[\tfrac12 k \text{ times}\big];0\Big)\\
&\qquad \times \Big\{\mu\Big(2,2,\dots\big[\tfrac12 k \text{ times}\big];0\Big) - \nu_2^{k/2}\Big\}\\
&\qquad + D\Big(2,2,\dots\big[\tfrac12 (k-2) \text{ times}\big];2\Big)W\Big(2,2,\dots\big[\tfrac12 (k-2) \text{ times}\big];2\Big)\\
&\qquad \times \mu\Big(2,2,\dots\big[\tfrac12 (k-2) \text{ times}\big];2\Big) + O\big(n^{-3/2}\big).
\end{aligned}
$$

Observe that for even k,

$$
\begin{aligned}
D\Big(2,2,\dots\big[\tfrac12 k \text{ times}\big];0\Big) &= k!\Big/\Big\{\big(\tfrac12 k\big)!2^{k/2}\Big\},\\
D\Big(2,2,\dots\big[\tfrac12 (k-2) \text{ times}\big];2\Big) &= k!k\Big/\Big\{\big(\tfrac12 k\big)!2^{(k+1)/2}\Big\},\\
n^{-k/2}W\Big(2,2,\dots\big[\tfrac12 (k-j) \text{ times}\big];j\Big) - (-1)^{j/2}\overline{w}_2^{k/2} &= O\big(n^{-1}\big)
\end{aligned}
$$

for $j = 0$ and 2, and

$$
\begin{aligned}
\mu\Big(2,2,\dots\big[\tfrac12 k \text{ times}\big];0\Big) &= \nu_2^{k/2} + n^{-1}\alpha_{k/2,1},\\
\mu\Big(2,2,\dots\big[\tfrac12 (k-2) \text{ times}\big];2\Big) &= n^{-1}\alpha_{k/2,2}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&E\Big\{\big(n^{1/2}S_1\big)^k\Big\} - E\Big\{\big(n^{1/2}S_{01}\big)^k\Big\}\\
&\quad = I(k \text{ even})n^{-1}\overline{w}_2^{k/2}\frac{k!}{(1/2k)!2^{k/2}}\big(\alpha_{k/2,1} - \tfrac12 k\alpha_{k/2,2}\big) + O\big(n^{-3/2}\big).
\end{aligned}
\tag{3.3}
$$

A slightly longer argument shows that when k is even, the remainder $O(n^{-3/2})$ may be replaced by $O(n^{-2})$. This proves (i)(a).

Next we establish (i)(b). Observe that

$$
\begin{aligned}
\sigma\tau^{-1} &= \Big\{1 + \sigma^{-2}\big(S_2 - 2\overline{v}_1 S_1 - S_1^2\big)\Big\}^{-1/2}\\
&= 1 - \tfrac12\sigma^{-2}\big(S_2 - 2\overline{v}_1 S_1 - S_1^2\big) + \tfrac38\sigma^{-4}(S_2 - 2\overline{v}_1 S_1)^2 + O_p\big(n^{-3/2}\big),
\end{aligned}
$$

whence $T \equiv \sigma\tau^{-1}S_1 = T_1 + O_p(n^{-3/2})$. Under the stated conditions on the sequence $\{N_i\}$, it may be proved that

$$E(T_1^k) \tag{3.4}$$
$$= \begin{cases} E\Big[S_1^k\{1 - \frac{1}{2}k\sigma^{-2}(S_2 - 2\bar{v}_1S_1)\}\Big] + O_p\big(n^{-(k+3)/2}\big), & \text{if } k \text{ is odd,} \\ E\Big[S_1^k\{1 - \frac{1}{2}k\sigma^{-2}(S_2 - 2\bar{v}_1S_1) + \frac{1}{2}k\sigma^{-2}S_1^2 \\ \qquad + \frac{1}{8}k(k+2)\sigma^{-4}(S_2 - 2\bar{v}_1S_1)^2\}\Big] + O_p\big(n^{-(k+4)/2}\big), & \text{if } k \text{ is even,} \end{cases}$$

and that an identical formula holds if (S, S_1, T_1) is replaced by (S_0, S_{01}, T_{01}) throughout. Arguments similar to those leading to (3.3) show that

$$n^{(k+1)/2}\Big[E\{S_1^k(S_2 - 2\bar{v}_1S_1)\} - E\{S_{01}^k(S_{02} - 2\bar{v}_1S_{01})\}\Big] = \begin{cases} O(n^{-1}), & \text{if } k \text{ is odd,} \\ O(n^{-3/2}), & \text{if } k \text{ is even,} \end{cases}$$

and that with $U = S_1$ or $S_2 - 2\bar{v}_1S_1$ and $U_0 = S_{01}$ or $S_{02} - 2\bar{v}_1S_{01}$, respectively,

$$n^{(k+2)/2}\Big\{E(S_1^kU^2) - E(S_{01}^kU_0^2)\Big\} = \begin{cases} O(n^{-1}), & \text{if } k \text{ is odd,} \\ O(n^{-3/2}), & \text{if } k \text{ is even.} \end{cases}$$

Result (i)(b) follows from these formulae and (i)(a). □

PROOF OF THEOREM 2.2. Observe that $E(S_{01}) = 0$, $E(S_{01}^2) = n^{-1}\nu_2t^2$, $E(S_{01}^3) = n^{-2}\nu_3\bar{w}_3$,

$$\begin{aligned} E(S_{01}^4) &= n^{-2}3\nu_2^2t^4 + n^{-3}(\nu_4 - 3\nu_2^2)\bar{w}_4 + O(n^{-4}), \\ E(S_{01}^5) &= n^{-3}10\nu_2\nu_3\bar{w}_2\bar{w}_3 + O(n^{-4}), \\ E(S_{01}^6) &= n^{-3}15\nu_2^3\bar{w}_2^3 + O(n^{-4}). \end{aligned}$$

The cumulants of S_{01} follow from these formulae.

Next, note that

$$\begin{aligned} E\{S_{01}(S_{02} - 2\bar{v}_1S_{01})\} &= n^{-1}\nu_2\bar{w}_3 + O(n^{-2}), \\ E\{S_{01}^2(S_{02} - 2\bar{v}_1S_{01})\} &= n^{-2}\nu_3(\bar{w}_4 - \bar{w}_2^2) + O(n^{-3}), \\ E\{S_{01}^3(S_{02} - 2\bar{v}_1S_{01})\} &= n^{-2}3\nu_2^2\bar{w}_2\bar{w}_3 + O(n^{-3}), \\ E\{S_{01}^4(S_{02} - 2\bar{v}_1S_{01})\} &= n^{-3}\nu_2\nu_3\Big\{4\bar{w}_3^2 + 6\bar{w}_2(\bar{w}_4 - \bar{w}_2^2)\Big\} + O(n^{-4}), \\ E\{S_{01}^2(S_{02} - 2\bar{v}_1S_{01})\} &= n^{-2}\nu_2^2\Big\{\bar{w}_2(\bar{w}_4 - \bar{w}_2^2) + 2\bar{w}_3^2\Big\} + O(n^{-3}), \\ E\{S_{01}^4(S_{02} - 2\bar{v}_1S_{01})^2\} &= n^{-3}\nu_2^3\Big\{3\bar{w}_2^2(\bar{w}_4 - \bar{w}_2^2) + 12\bar{w}_2\bar{w}_3^2\Big\} + O(n^{-4}). \end{aligned}$$

Therefore, by (3.4),

$$
\begin{aligned}
E(T_{01}) &= -\tfrac{1}{2}n^{-1}\nu_2 t^{-2}\overline{w}_3 + O(n^{-2}),\\
E(T_{01}^2) &= n^{-1}\nu_2 t^{-2} + n^{-2}\Big\{-\nu_3 t^{-2}(\overline{w}_4 - t^4) + 3\nu_2^2 t^2\\
&\qquad + \nu_2^2\big(t^{-2}\overline{w}_4 - t^2 + 2t^{-4}\overline{w}_3^2\big)\Big\} + O(n^{-3}),\\
E(T_{01}^3) &= n^{-2}\big(\nu_2\overline{w}_3 - \tfrac{9}{2}\nu_2^2\overline{w}_3\big) + O(n^{-3}),\\
E(T_{01}^4) &= n^{-2}3\nu_2^2 t^4 + n^{-3}\Big[(\nu_4 - 3\nu_2^2)\overline{w}_4\\
&\qquad - 4\nu_2\nu_3 t^{-2}\Big\{2\overline{w}_3^2 + 3t^2(\overline{w}_4 - t^4)\Big\} + 30\nu_2^3 t^4\\
&\qquad + 9\nu_2^3 t^{-2}\Big\{t^2(\overline{w}_4 - t^4) + 4\overline{w}_3^2\Big\}\Big] + O(n^{-4}).
\end{aligned}
$$

The cumulants of T_{01} follow from these formulae. □

Acknowledgments. We are grateful to three referees and the editors for their constructive and detailed comments on an earlier version.

REFERENCES

BARBE, P. and BERTRAIL, P. (1993). *The Weighted Bootstrap*. Série Document de Travail: Méthodologie Statistique. Institut National de la Recherche Agronomique Secteur des Sciences Sociales, Ivry Sur Seine.

BERAN, R. (1986). Discussion of "Jackknife, bootstrap and other resampling methods in regression analysis," by C. F. J. Wu. *Ann. Statist.* **14** 1295–1298.

BICKEL, P. J. and FREEDMAN, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9** 1196–1217.

EFRON, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist.* **7** 1–26.

HAEUSLER, E., MASON, D. M. and NEWTON, M. A. (1992). Weighted bootstrapping of means. *CWI Quarterly* **5** 213–228.

HALL, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer, New York.

HÄRDLE, W. and MAMMEN, E. (1990). Comparing nonparametric versus parametric regression fits. Preprint. Sonderforschungsbereich 123, Univ. Heidelberg.

HUŠKOVÁ, M. and JANSSEN, P. (1993a). Generalized bootstrap for studentized U-statistics: A rank statistic approach. *Statist. Probab. Lett.* **16** 225–233.

HUŠKOVÁ, M. and JANSSEN, P. (1993b). Consistency of the generalized bootstrap for degenerate U-statistics. *Ann. Statist.* **21** 1811–1823.

KENDALL, M. G., STUART, A. and ORD, J. K. (1987). *Kendall's Advanced Theory of Statistics. Distribution Theory* **1**. Oxford Univ. Press.

LIU, R. (1988). Bootstrap procedures under some non i.i.d. models. *Ann. Statist.* **16** 1696–1708.

LO, A. Y. (1987). A large sample study of the Bayesian bootstrap. *Ann. Statist.* **15** 360–375.

MAMMEN, E. (1992). *When Does Bootstrap Work: Asymptotic Results and Simulations. Lecture Notes in Statist*. **77**. Springer, New York.

MASON, D. M. and NEWTON, M. A. (1992). A rank statistics approach to the consistency of a general bootstrap. *Ann. Statist.* **20** 1611–1624.

PRAESTGAARD, J. and WELLNER, J. A. (1993). Exchangeably weighted bootstraps of the general empirical process. *Ann. Probab.* **21** 2053–2086.

RUBIN, D. B. (1981). The Bayesian bootstrap. *Ann. Statist.* **9** 130–134.

SILVERMAN, B. W. and YOUNG, G. A. (1987). The bootstrap: to smooth or not to smooth? *Biometrika* **74** 469–479.

TAYLOR, R. L., DAFFER, P. Z. and PATTERSON, R. F. (1985). *Limit Theorems for Sums of Exchangeable Random Variables*. Rowman and Allenheld, Totowa, NJ.

WENG, C.-S. (1989). On a second-order asymptotic property of the Bayesian bootstrap mean. *Ann. Statist.* **17** 705–710.

WU, C. F. J. (1986). Jackknife, bootstrap, and other resampling methods in regression analysis (with discussion). *Ann. Statist.* **14** 1261–1295.

YOUNG, G. A. (1988). A note on bootstrapping the correlation coefficient. *Biometrika* **75** 370–373.

CENTRE FOR MATHEMATICS AND
ITS APPLICATIONS
AUSTRALIAN NATIONAL UNIVERSITY
CANBERRA ACT 0200
AUSTRALIA

INSTITUT FÜR ANGEWANDTE MATHEMATIK
UNIVERSITÄT HEIDELBERG
HEIDELBERG
GERMANY