The Annals of Statistics
1998, Vol. 26, No. 6, 2369–2394

BOUNDS FOR PROBABILITIES OF SMALL RELATIVE ERRORS FOR EMPIRICAL SADDLEPOINT AND BOOTSTRAP TAIL APPROXIMATIONS

BY J. ROBINSON AND IB M. SKOVGAARD[1]

University of Sydney and Royal Veterinary and Agricultural University

To obtain test probabilities based on empirical approximations to the distribution of a Studentized function of a mean, we need the approximations to be accurate with sufficiently high probability. In particular, when these test probabilities are small it is best to consider relative errors. Here we show that in the case of univariate standardized means and in the general case of tests based on smooth functions of means, the empirical approximations have asymptotically small relative errors on sets with probability differing from 1 by an exponentially small quantity and that these error rates hold for moderately large deviations. In particular, for standardized deviations of order $n^{1/6}$, the probabilities approximated are exponentially small with exponents of order $n^{1/3}$ and the corresponding relative errors tend to zero on sets whose complements have probabilities of the order of the probabilities being approximated.

1. Introduction. Consider samples taken from a distribution of a random variable X with $EX = \mu$ and $\mathrm{var}(X) = \sigma^2$. To obtain p-values for tests of hypothesis on μ when σ^2 is known we need to calculate

$$\mathrm{P}\biggl(\frac{\bar{X}-\mu}{\sigma} \geq a\biggr).$$

If a moment generating function of X exists, then this tail probability can be approximated using saddlepoint methods. Under regularity conditions the following result holds for $|a| < C$:

$$\mathrm{P}\biggl(\frac{\bar{X}-\mu}{\sigma} \geq a\biggr) = Q(a)\biggl(1 + O\biggl(\frac{1}{n}\biggr)\biggr), \tag{1.1}$$

where $Q(a)$ will be defined later.

If the distribution, and so the cumulant generating function, of X is unknown, then we can attempt to approximate the tail probability by an empirical saddlepoint as in Jing, Feuerverger and Robinson (1994), extending the idea first proposed by Feuerverger (1989). Here we consider a sample $X_1, \dots, X_n$ and use the sample mean $\bar{X}$, the sample variance S^2 and the empirical cumulant generating function. Then the empirical saddlepoint approximation to the tail probability will be written $\hat{Q}(a)$.

Received June 1997; revised August 1998.
[1]Supported for a period at the University of Sydney by an A.R.C. Institutional Grant.
AMS 1991 *subject classifications*. Primary 62E20, 60F05.
Key words and phrases. Large deviations, smooth functions of means, exponential error rates.

In more general cases where X is a d-dimensional random vector, we consider a test statistic $g(\bar{X})$ to test some hypothesis $g(\mu) = 0$, where the function g may depend on $EX = \mu$ and $\operatorname{cov}(X) = \Sigma$. To obtain p-values we need to calculate $\mathrm{P}(g(\bar{X}) \geq a)$. Again, if a moment generating function of X exists then this tail probability can be approximated using saddlepoint methods. For $|a| < C$ we have

$$\mathrm{P}\big(g(\bar{X}) \geq a\big) = Q(a)\left(1 + O\left(\frac{1}{n}\right)\right), \tag{1.2}$$

where again $Q(a)$ will be defined later. Again the empirical saddlepoint approximation will be written $\hat{Q}(a)$.

In both these cases, a bootstrap approximation to the probabilities could be used in place of the empirical saddlepoint. In fact, it can be seen that the empirical saddlepoint is just the saddlepoint approximation to the bootstrap approximation. Saddlepoint approximations to the bootstrap were considered in Davison and Hinkley (1988), Daniels and Young (1991) and DiCiccio, Martin and Young (1992), in which papers there are a number of numerical illustrations of the accuracy of the saddlepoint approximation. It is of interest to consider the relative errors of these approximations.

In Jing, Feuerverger and Robinson (1994), bounds in probability were found for the relative error of these approximations. Indeed it was shown that the relative error was $O_p((1 + \sqrt{n}a)^3/n)$ for $\sqrt{n}a = o(n^{1/3})$. However, it is of interest to give a rate of convergence for the probability that the relative error exceeds such a bound. It is of importance to know the relationship of this rate to the size of the tail area probability which is being approximated.

In the present paper we consider the standardized variable $0 < \sqrt{n}a \leq \eta n^{\beta}$, so the tail probability at the extreme of this interval is of order $\exp(-cn^{2\beta})$, for $\beta > 0$ sufficiently small. It is proved that for $0 \leq \beta < \frac{1}{3}$ we can choose $0 < \delta \leq 1 - 3\beta$ and $0 < \alpha < \min\{\frac{1}{6} + \delta/3, 2\delta\}$ so that with probability $1 - \exp(-cn^{\alpha})$, the empirical saddlepoint and the bootstrap approximations to the tail probability both have relative error bounded by a quantity of order $(1 + \sqrt{n}a)^3 n^{-1+\delta}$. If β is close to $1/3$ and so δ and α are close to 0, then the relative error is bounded by $O((1 + \sqrt{n}a)^3 n^{-1})$ with probability of $o(1)$ corresponding to the results of Jing, Feuerverger and Robinson (1994). If $\beta = 1/6$ we can take $\delta = 1/2$ and α arbitrarily close to $1/3$. In this case the bound for the probability of the relative error exceeding $O((1 + \sqrt{n}a)^3 n^{-1/2})$ matches the tail probability. For $\frac{1}{6} < \beta < \frac{1}{3}$ we cannot obtain a probability of the relative error exceeding the bound as small as the tail probability, while, for $0 \leq \beta < \frac{1}{6}$, we can choose $0 < \delta \leq 1 - 3\beta$ and take α large enough to obtain smaller bounds on the relative errors with probability of the relative errors exceeding these bounds as small as the tail probability. In the limiting case with $\beta = 0$, we can take δ and α arbitrarily close to 0 achieving a relative error of almost n^{-1} with small probability. This corresponds to the usual second order correctness of the absolute error.

This is the equivalent in the relative error case of results such as that in Theorem 5.1 of Hall (1992), which is used to show that the probability of the bootstrap approximation having an absolute error of more than $O(n^{-1})$ is $O(n^{-1-\varepsilon})$. Results of this type were considered for permutation tests in Albers, Bickel and van Zwet (1976) and Bickel and van Zwet (1978). Hall (1990) considers large deviation results for the bootstrap but does not obtain bounds for the probability of relative errors as discussed here.

In Section 2 we give the notation and the main results on the empirical saddlepoint approximation in the form of two theorems, the first giving the result for a univariate standardized mean and the second for the general case. A detailed proof is given for the first case based on a technical lemma concerning rates for tilted cumulants, whose proof is deferred to Section 4. It is shown that the proof for the general case follows exactly the lines of that for the univariate case. In Section 3 we consider the bootstrap approximation by comparing it to the empirical saddlepoint. It is clear from this comparison that the critical errors in the bootstrap approximation are in the relative errors of the empirical saddlepoint to the true probabilities rather than in the relative errors of the bootstrap and the empirical saddlepoint. In Section 4 we obtain a number of new bounds for moments and cumulants which are of independent interest.

2. Main results.

This section contains the main results concerning the relative error of the empirical saddlepoint approximation to the tail probabilities for the univariate standardized mean (Theorem 1), and for the multivariate case (Theorem 2) which covers the univariate Studentized mean as a special case.

2.1. *Notation.* Let $X_1, \dots, X_n$ be independent identically distributed random variables with $EX_1 = \mu$ and $\operatorname{var} X_1 = \sigma^2 > 0$. Assume that $K(t) = \log \mathrm{E}\exp(tX_1)$ is bounded in $|t| < A$ for some $A > 0$. For convenience let $a > 0$; then the saddlepoint approximation for the tail probability in (1.1) is

$$Q(a) = 1 - \Phi\big(\sqrt{n}w^*(a)\big), \tag{2.1}$$

where

$$w^*(a) = w(a) - \frac{\log \psi(a)}{nw(a)},$$

for $w(a) = \sqrt{2H(a)}$, where

$$\Lambda(x) = \inf_t \big\{tx - K(t)\big\},$$

$$H(a) = \Lambda(\mu + \sigma a) = \inf_x \left\{\Lambda(x)\colon \frac{x-\mu}{\sigma} = a\right\}, \tag{2.2}$$

where $\mu = K'(0)$ and $\sigma^2 = K''(0)$, and for

$$\psi(a) = \frac{w(a)}{t(a)\sqrt{K''(t(a))}},$$

where $t(a)$ is the value minimizing $tx - K(t)$ for $x = x(a) = \mu + \sigma a$. We will refer to this as the Barndorff-Nielsen approximation following his r^* statistic in Barndorff-Nielsen (1986). Its derivation is in Jing, Feuerverger and Robinson (1994) and it is shown to be equivalent to the Lugannani–Rice approximation in Jensen (1992).

The empirical saddlepoint approximation is obtained by replacing $K(t)$ by

$$\hat{K}(t) = \log \frac{1}{n} \sum_{j=1}^{n} \exp(tX_j);$$

then the approximation is given by

$$\hat{Q}(a) = 1 - \Phi(\sqrt{n}\hat{w}^*(a)),$$

where $\hat{w}^*$ is defined in the same way as w^* with $\hat{K}$ replacing K. Note that μ is replaced by $\bar{X} = \hat{K}'(0)$, σ^2 by $\hat{\sigma}^2 = \hat{K}''(0)$, $t(a)$ by $\hat{t}(a)$ and $H(a)$ by $\hat{H}(a)$.

For the general case, consider independent identically distributed random vectors $X_1, \dots, X_n \in \mathbf{R}^d$ with cumulant generating function $K(t) = \log \mathrm{E} \exp(t^T X_1)$ assumed to be finite for $\|t\| < A$, $t \in \mathbf{R}^d$. Let $\mu = \mathrm{E} X_1 = K'(0)$ denote the mean vector, and $\Sigma = \operatorname{var} X_1 = K''(0)$ the variance matrix.

We consider tail probabilities

$$\mathrm{P}\big(g(\bar{X}; \mu, \Sigma) \geq a\big), \tag{2.3}$$

still for $a > 0$, where g is a smooth one-dimensional function satisfying certain conditions ensuring that $\sqrt{n}g(\bar{X}; \mu, \Sigma)$ is asymptotically standard normal. In the sequel we use the abbreviations $g(x) = g(x; \mu, \Sigma)$ and $\hat{g}(x) = g(x; \hat{\mu}, \hat{\Sigma})$, unless the functional dependence on the two moments is of direct concern. The setting is designed to cover Studentized statistics such as

$$\big(f(\bar{Y}) - f(\mu_Y)\big)/\big\{f'(\mu_Y)^T \hat{V} f'(\mu_Y)\big\}^{1/2},$$

where $Y_1, \dots, Y_n \in \mathbf{R}^q$ are independent and identically distributed, $\mu_Y = \mathrm{E}Y_1$, f is a smooth real function, and

$$\hat{V} = \frac{1}{n} \sum_{1}^{n} Y_i Y_i^T - \bar{Y}\bar{Y}^T$$

is the empirical variance matrix. In this case we should take $X_i = (Y_i, Y_i Y_i^T)$; see Hall [(1992), Section 2.4]. A standardized statistic uses $\operatorname{var} Y_1$, which is a subblock of Σ, instead of $\hat{V}$, and another variant of the Studentized statistic is obtained by use of $f'(\bar{Y})$ instead of $f'(\mu_Y)$.

In any case we require that the function g satisfies the conditions

$$g(\mu; \mu, \Sigma) = 0 \tag{2.4}$$

and

$$g'(\mu; \mu, \Sigma)^T \Sigma g'(\mu; \mu, \Sigma) = 1 \tag{2.5}$$

for any μ and Σ, where g' refers to the derivative with respect to the first argument of g. As in the one-dimensional case, the saddlepoint approximation

$Q(a)$ is defined in (2.1), using the previous definitions of $w^*(a)$ and $w(a)$, except that $H(a)$ and $\psi(a)$ are replaced by their generalizations. Thus, $H(a)$ is now defined by the equations

$$\begin{aligned}\Lambda(x) &= \inf_t \{t^T x - K(t)\},\\ H(a) &= \Lambda(x(a)) = \inf_x \{\Lambda(x)\colon g(x) = a\}.\end{aligned} \tag{2.6}$$

We let $x(a)$ denote the vector minimizing the second equation and $t(a)$ the vector minimizing the first for $x = x(a)$. From Jing and Robinson (1994), Theorem 3 and equations (4.2) and (4.4), we have

$$\psi(a) = w(a)G(a)/H'(a), \tag{2.7}$$

where, using x for $x(a)$, t for $t(a)$, I_d for the $d \times d$ identity matrix and $|\cdot|$ to denote a determinant,

$$G(a) = \left|I_d - H'(a)K''(t)g''(x)\right|^{-1/2}\left|g'(x)^T K''(t)g'(x)\right|^{-1/2} \tag{2.8}$$

because, in our case, $H'(a)$ equals their Lagrangian multiplier λ. Related results appear in DiCiccio, Field and Fraser (1990) and Daniels and Young (1991).

The empirical saddlepoint approximation $\hat{Q}(a)$ is defined in the same way, using $\hat{K}(t)$ instead of $K(t)$, $\hat{\mu} = \hat{K}'(0)$ instead of μ and $\hat{\Sigma} = \hat{K}''(0)$ instead of Σ.

Positive constants like c, C and η may differ from occurrence to occurrence.

2.2. *Univariate results.* Here we will prove the following result.

THEOREM 1. *For $0 \le \beta < \frac{1}{3}$, and for any $c > 0$ and $C > 0$, there exist positive constants η and C' such that for $0 < \delta \le 1 - 3\beta$ and $0 < \alpha < \min\{\frac{1}{6} + \delta/3, 2\delta\}$ we have, for large enough n,*

$$\mathrm{P}\left(\frac{|\hat{Q}(a) - Q(a)|}{Q(a)} \le C(1 + \sqrt{n}a)^3 n^{-1+\delta}, 0 < a < \eta n^{\beta - 1/2}\right) > 1 - C'\exp(-cn^{\alpha}).$$

The proof is based on inequalities for moments given by the following lemma, the proof of which, together with a number of preliminary results on moments, is given in Section 4.

LEMMA 1. *For any $0 < \alpha < 1$, $L \in \mathbf{N}$, $C > 0$ and $c > 0$, there exist constants $C' > 0$ and $\eta' > 0$, such that*

$$\mathrm{P}\left(|\hat{K}^{(k)}(t) - K^{(k)}(t)| < Cn^{-1/2+\delta_k}, \qquad |t| < \eta' n^{-\alpha}\right) > 1 - C'\exp(-cn^{\alpha}), \tag{2.9}$$

for $k = 0, 1, \dots, L$, and any δ_k's satisfying

$$\delta_k - \tfrac{1}{2}\alpha > 0 \quad \textit{and} \quad \tfrac{1}{2} - k\alpha + \delta_k > 0 \tag{2.10}$$

for $k = 1, \dots, L$ and $\delta_0 = \delta_1 - \alpha$.

REMARK. Note that under the conditions of the Theorem the restriction $|t| < \eta' n^{-\alpha}$ in (2.9) can be replaced by $|t| < \eta' n^{\beta-1/2}$ since this is a more restricted range.

PROOF OF THEOREM 1. Noting that $\phi(x)/\{1-\Phi(x)\} < 1+x,\ x > 0$, which implies that for u, v positive,

$$\frac{|\Phi(v)-\Phi(u)|}{1-\Phi(u)} \le (1+u)|u-v|\exp(u|u-v|), \tag{2.11}$$

we can show

$$\begin{aligned}\frac{|\hat{Q}(a)-Q(a)|}{Q(a)} &= \frac{|\Phi(\sqrt{n}\hat{w}^*(a))-\Phi(\sqrt{n}w^*(a))|}{1-\Phi(\sqrt{n}w^*(a))}\\ &\le C''\sqrt{n}\big|\hat{w}^*(a)-w^*(a)\big|\big(1+\sqrt{n}w^*(a)\big)\end{aligned} \tag{2.12}$$

for $\sqrt{n}|\hat{w}^*(a)-w^*(a)|\,|1+\sqrt{n}w^*(a)|$ sufficiently small.

Assume that the inequalities stated in Lemma 1 are satisfied; this is true with the desired probability. For application of the lemma, the δ_k's should be chosen appropriately small, in particular as an increasing sequence with $\delta_2 < 1/2$. To bound differences, like $\hat{w}^*(a) - w^*(a)$, between empirical quantities and their theoretical analogues we need first to show that $|\hat{t}(a)| < c' n^{\beta-1/2}$ for suitable $c' > 0$, so that we may choose η such that $\hat{t}(a)$ is within the range required for the inequalities in Lemma 1 to hold.

For any u such that $|\hat{t}(u)| < c' n^{\beta-1/2}$ we then have

$$\hat{t}'(u) = \frac{\hat{K}''(0)^{1/2}}{\hat{K}''(\hat{t}(u))} = \frac{1}{\sigma} + \theta\big(n^{\delta_2-1/2} + n^{\beta-1/2}\big), \tag{2.13}$$

where θ is bounded by some constant, and we have used $|K''(t(u)) - K''(0)| < C''|t(u)|$ is of order $n^{\beta-1/2}$.

Recall that $\hat{t}(0) = 0$ and let

$$\bar{a} = \inf_a \big\{a > 0\colon |\hat{t}(a)| \ge c' n^{\beta-1/2}\big\}.$$

If no such a exists, $\hat{t}(a)$ is within the range given in the lemma for all $a > 0$. Now, by the mean value theorem,

$$\hat{t}(\bar{a}) = \hat{t}'(u)\bar{a}$$

for some u between 0 and $\bar{a}$ and hence with $|\hat{t}(u)| < c' n^{\beta-1/2}$. Thus, for n sufficiently large,

$$\bar{a} = \frac{|\hat{t}(\bar{a})|}{|\hat{t}'(u)|} > c' n^{\beta-1/2}\frac{\sigma}{2},$$

say.

Since this argument is valid for any sufficiently small $c' > 0$ we have shown that for any $c' > 0$ we may choose $c'' > 0$ such that on a set of probability $1 - C' \exp(-cn^{\alpha})$,

$$a < c'' n^{\beta - 1/2} \Rightarrow |\hat{t}(a)| < c' n^{\beta - 1/2}. \tag{2.14}$$

Now we may assume that η' has been chosen such that $\hat{t}(a) < \eta' n^{\beta-1/2} < \eta' n^{-\alpha}$.

Returning to (2.12) and the inequality

$$\left|\hat{w}^*(a) - w^*(a)\right| \le \left|\hat{w}(a) - w(a)\right| + \frac{1}{n}\left|\frac{\log \hat{\psi}(a)}{\hat{w}(a)} - \frac{\log \psi(a)}{w(a)}\right|, \tag{2.15}$$

we now consider the first difference on the right for which we have

$$\begin{aligned} \left|\hat{w}(a) - w(a)\right| &= a\left|\sqrt{2\hat{H}(a)/a^2} - \sqrt{2H(a)/a^2}\right| \\ &\le a\frac{\left|\hat{H}(a) - H(a)\right|}{a^2} \end{aligned} \tag{2.16}$$

for a and $|\hat{H}(a) - H(a)|/a^2$ sufficiently small, which will be the case since $H(0) = 0$, $H'(0) = 0$ and $H''(0) = 1$, as shown in Jing, Feuerverger and Robinson (1994).

To bound differences of the form

$$\frac{|\hat{f}(a) - f(a)|}{a},$$

where f is some function with $\hat{f}(0) = f(0)$, we use

$$\frac{1}{a}\left|(\hat{f} - f)(a)\right| \le \sup_{0<u<a}\left\{|\hat{f}'(u) - f'(u)|\right\}. \tag{2.17}$$

If also $\hat{f}'(0) = f'(0)$, then the same method gives

$$\frac{1}{a}\left|(\hat{f} - f)(a)\right| \le \frac{a}{2}\sup_{0<u<a}\left\{|\hat{f}''(u) - f''(u)|\right\}, \tag{2.18}$$

and so on.

In this way we obtain

$$\frac{|\hat{H}(a) - H(a)|}{a^2} = \frac{a}{6}\sup_{0<u<a}\left\{|\hat{H}^{(3)}(u) - H^{(3)}(u)|\right\}. \tag{2.19}$$

Now

$$H^{(3)}(a) = K^{(3)}(t(a))K''(0)^{3/2}K''(t(a))^{-3}$$

with an equivalent result for $\hat{H}$. So we can bound $|\hat{H}^{(3)}(u) - H^{(3)}(u)|$ using

$$|\hat{A}\hat{B}\hat{C} - ABC| \le |\hat{A} - A|\,|BC| + |\hat{A}|\,|\hat{B} - B|\,|C| + |\hat{A}\hat{B}|\,|\hat{C} - C|. \tag{2.20}$$

For the first difference in (2.20) we now use Lemma 1 to obtain

$$\begin{aligned}
&\sup_{0<a<\eta n^{\beta-1/2}} \left|\hat{K}^{(3)}(\hat{t}(a)) - K^{(3)}(t(a))\right| \\
&\quad \le \sup_{0<a<\eta n^{\beta-1/2}} \left|\hat{K}^{(3)}(\hat{t}(a)) - K^{(3)}(\hat{t}(a))\right| \\
&\qquad + \sup_{0<a<\eta n^{\beta-1/2}} |\hat{t}(a) - t(a)| \sup_{0<a<\eta n^{\beta-1/2}} K^{(4)}(t(a)) \\
&\quad < C_1 n^{-1/2+\delta_3}
\end{aligned} \tag{2.21}$$

with the desired probability, for some $C_1 > 0$, where the last inequality uses

$$|\hat{t}(a) - t(a)| \le a \sup_{0<u<a} |\hat{t}'(u) - t'(u)| < C'' n^{\beta-1/2} n^{\delta_2-1/2} < C'' n^{\delta_3-1/2} \tag{2.22}$$

for some C''.

A similar argument can be used for $\hat{K}''(\hat{t}(a)) - K''(t(a))$ and $\hat{K}''(0) - K''(0)$, then from (2.20),

$$\mathrm{P}\left(|\hat{H}^{(3)}(a) - H^{(3)}(a)| < C_2 n^{-1/2+\delta_3}, 0 < a < \eta n^{\beta-1/2}\right) > 1 - C' \exp(-cn^{\alpha}) \tag{2.23}$$

for some $C_2 > 0$. Tracing the constants through the argument it may be seen that (2.23) holds for any $C_2 > 0$ if other constants are properly chosen.

Applying (2.19) and (2.23) in (2.16), we have,

$$\mathrm{P}\left(|\hat{w}(a) - w(a)| < C_2 a^2 n^{-1/2+\delta_3}, 0 < a < \eta n^{\beta-1/2}\right) > 1 - C' \exp(-cn^{\alpha}).$$

For the second difference on the right in (2.15) we may write

$$\frac{\log \psi(a)}{w(a)} = \frac{a}{w(a)} \frac{\frac{1}{2}\log\{2H(a)/a^2\} - \log\{H'(a)/a\} + \frac{1}{2}\log H''(a)}{a}, \tag{2.24}$$

consider the three terms in the numerator of the last fraction one by one and note that the factor $a/w(a)$ in front gives rise only to an unimportant relative error compared to $(\log \psi(a))/w(a)$. For each of the three terms, the difference between the empirical term and its theoretical counterpart is bounded by use of (2.17) or (2.18) in the same way as the bound for $|\hat{w}(a) - w(a)|$ was obtained. For each of the three differences the critical estimation error is for $\sup_u\{|\hat{H}^{(3)}(u) - H^{(3)}(u)|\}$ which is of order $n^{-1/2+\delta_3}$ with the stated probability.

Also, $w(a)/a$ and $\log \psi(a)/w(a)$ are bounded when a is bounded, so using these results in (2.12) we have the result. □

2.3. *Multivariate results.* For the general case we have the following result which gives the same rates as for the univariate case.

THEOREM 2. *Assume that $g(\bar{X}; \mu, \Sigma)$ is three times continuously differentiable in $\bar{X}$, and that these derivatives, as well as g itself, are continuously* *differentiable with respect to* μ *and* Σ *at the values corresponding to the distribution* P *of* X_1. *With* α, β, δ *and other constants as in Theorem* 1,

$$\mathrm{P}\left(\left|\frac{\hat{Q}(a)-Q(a)}{Q(a)}\right| \le C\big(1+\sqrt{n}a\big)^3 n^{-1+\delta},\, 0<a<\eta n^{\beta-1/2}\right) > 1-C'\exp(-cn^{\alpha}). \tag{2.25}$$

The proof follows the lines of the univariate case, but needs the following generalization of Lemma 1 to higher dimension. The proof of Lemma 2 is given in Section 4.2.

LEMMA 2. *Let*

$$K_{i_1\cdots i_k}(t) = \frac{\partial^k K(t)}{\partial t_{i_1}\cdots\partial t_{i_k}}$$

denote the kth partial derivative of K *at* t *with respect to* k *coordinates of* t. *Then, for any* $L \in \mathbf{N}$, $C>0$, $c>0$, $\alpha<1$, *there exist constants* $\eta'>0$ *and* $C'>0$ *such that*

$$\mathrm{P}\big(\big|\hat{K}_{i_1\cdots i_k}(t) - K_{i_1\cdots i_k}(t)\big| < Cn^{-1/2+\delta_k},\, |t|<\eta' n^{-\alpha}\big) > 1-C'\exp(-cn^{\alpha}), \tag{2.26}$$

for all $k=0,\dots,L$, *and* $i_1,\dots,i_k$, *where the* δ_k*'s satisfy the same condition as in Lemma* 1.

REMARK. Bounding the differences of the partial derivatives in Lemma 2 is equivalent to bounding the norms of the differences of the differentials considered as multidimensional objects.

PROOF OF THEOREM 2. We start from inequality (2.12) and continue as in the univariate case to inequality (2.15) for which similar computations will give the same result as in the univariate case. In the sequel we concentrate on the alterations to the functions involved. Throughout the proof we assume that the event stated in Lemma 2 has occurred, which is the case with the desired probability.

As in the univariate case, it follows that $H(0)=0$, $H'(0)=0$, $H''(0)=1$, $w(0)=0$, $w'(0)=1$ and $\psi(0)=1$. Thus, we get the bound (2.16) for $|\hat{w}(a)-w(a)|$, if the two factors on the right tend to zero. For this and other similar inequalities we need the bounds $\|\hat{t}(a)\| < c'n^{\beta-1/2}$ and $\|\hat{t}(a)-t(a)\| < C''n^{\beta-1+\delta_2}$, which are obtained as in the univariate case, except that the expression for $\hat{t}'(a)$ is now slightly more complicated.

A bound for $|\hat{w}-w|$ is obtained as in (2.16), and for the ψ-part of $w^*(a)$ we have

$$\frac{\log\psi(a)}{w(a)} = \frac{a}{w(a)}\frac{\log\{w(a)/a\}-\log\{H'(a)/a\}+\log G(a)}{a}, \tag{2.27}$$

which corresponds closely to (2.24). The difference between each of the three terms in the last fraction and its empirical analogue is bounded again by use of (2.17) and (2.18). The first term is already taken care of; the third gives

$$\frac{1}{a}\left|\log \hat{G}(a) - \log G(a)\right| \leq C_1 \sup_u \left|(\log \hat{G})'(u) - (\log G)'(u)\right|$$

for some constant C_1 and the second gives

$$\frac{1}{a}\left|\frac{\log\{\hat{H}'(a)/a\} - \log\{H'(a)/a\}}{a}\right| \leq C_2 \sup_u \left|\hat{H}^{(3)}(u) - H^{(3)}(u)\right| \tag{2.28}$$

for some C_2, where we have bounded the logarithmic difference in terms of the difference itself, using the fact that the derivative of the logarithmic function is bounded in a neighborhood of $H'(a)/a$.

Thus, the crucial quantities are the suprema of differences between the empirical and theoretical versions of the third derivative of the function $H(a)$ and of the first derivative of $\log G(a)$. By use of a Lagrange multiplier to derive the infimum in the definition of $H(a)$, it may be shown that

$$H''(a) = \left\{g'(x)^T\left(I_d - H'(a)K''(t)g''(x)\right)^{-1}K''(t)g'(x)\right\}^{-1}.$$

For the differentiation of this expression we further need

$$x'(a) = H''(a)\left\{I_d - H'(a)K''(t)g''(x)\right\}^{-1}K''(t)g'(x),$$

and to make it explicit, also

$$H'(a)g'(x) = t.$$

By differentiation of the expression for $H''(a)$, we see that derivatives of $K(t)$ and $g(x)$ only up to third occur. All functions involved are bounded since the only matrices being inverted are the ones from the expression for $H''(a)$, which tend to I_d and 1, respectively, due to the assumptions on $g(x)$ and the range of a. It follows that the difference in (2.28) is bounded by the same rate as the $K^{(3)}(t)$ difference, which is known from Lemma 2 to be $n^{-1/2+\delta_3}$.

For the $\log G(a)$ difference, the same arguments may be applied, also leading to the $K^{(3)}(t)$ difference as the limiting quantity. A technical point here is that the first derivative of the logarithm of the determinant of a matrix M is given by the first order Taylor series approximation

$$\log|M + \Delta| = \log|M| + \operatorname{trace}(M^{-1}\Delta) + o(\|\Delta\|)$$

as $\|\Delta\| \to 0$. Since the determinant in (2.8) is of a matrix that tends to I_d, the inverse arising by differentiation causes no problem.

Thus, Lemma 2 ensures that we get the same result as for the univariate case, since the differences between empirical and theoretical third-order derivatives of the cumulant generating function are of order $n^{-1/2+\delta_3}$.

3. The bootstrap approximation.

The empirical saddlepoint of the previous section is just the saddlepoint approximation to the bootstrap. In this section we show that an appropriate relative error for this approximation holds except in sets with exponentially small probability.

3.1. *Notation and main result.* Consider the notation of Section 2.1. Let $Z_1, \dots, Z_n$ be a bootstrap sample from $X_1, \dots, X_n$. Write

$$\tilde{Q}(a) = \tilde{\mathrm{P}}\left(\frac{\bar{Z} - \bar{X}}{\hat{\sigma}} \geq a\right), \tag{3.1}$$

where $\tilde{\mathrm{P}}$ denotes the bootstrap probability conditional on $X_1, \dots, X_n$. Then the empirical saddlepoint approximation $\hat{Q}(a)$ is just the saddlepoint approximation to $\tilde{Q}(a)$.

Let E_n be the set of values of $X_1, \dots, X_n$ such that

$$\hat{K}^{(4)}(t)/\big[\hat{K}^{(2)}(t)\big]^2 < C\big(1 + n^{-1/2+\delta_4}\big) \quad \text{for } |t| < \eta' n^{\beta-1/2}, \tag{3.2}$$

$$\big|\exp(\hat{K}(t + i\xi) - \hat{K}(t))\big| < 1 - \varepsilon \quad \text{for } \gamma < |\xi| < \Gamma n,\ |t| < \eta' n^{\beta-1/2}. \tag{3.3}$$

Then $\{X_1, \dots, X_n\} \in E_n$ implies

$$\frac{|\tilde{Q}(a) - \hat{Q}(a)|}{\hat{Q}(a)} < \frac{C}{n}\left(\frac{|\hat{K}^{(4)}(\hat{t}(a))|}{\hat{K}^{(2)}(\hat{t}(a))^2} + C'\right) \quad \text{for } |a| < \eta n^{\beta-1/2}. \tag{3.4}$$

This result follows as a special case from Theorem 1 of Robinson, Höglund, Holst and Quine (1990) after approximating the indirect Edgeworth approximation given there by the Barndorff–Nielsen approximation as described in Jing and Robinson (1994).

THEOREM 3. *Assume that for $\gamma > 0$ there is a $d > 0$ such that*

$$|\phi(\xi)| = |Ee^{i\xi X}| < 1 - d \qquad \text{for all } |\xi| > \gamma. \tag{3.5}$$

For any constants $0 < \alpha < 1/2$ and $\delta_4 > 0$ satisfying $\alpha < \min\{1/8 + \delta_4/4, 2\delta_4\}$, then

$$\mathrm{P}\left(\frac{|\tilde{Q}(a) - \hat{Q}(a)|}{\hat{Q}(a)} \leq Cn^{-1}(1 + n^{-1/2+\delta_4}), 0 < a < \eta n^{-\alpha}\right) > 1 - C' \exp(-cn^{\alpha}).$$

REMARK. If we choose α as in Theorem 1, then for this result to be used in conjunction with the results of Theorem 1, we need to choose δ_4 as small as possible while keeping the conditions imposed in Theorem 1. Such a choice is always possible in such a way that $\tilde{Q}(a)$ can replace $\hat{Q}(a)$ in Theorem 1, showing that the bootstrap has the same properties as the empirical saddlepoint.

PROOF OF THEOREM 3. Assume, as in the proof of Theorem 1, that the inequalities of Lemma 1 are satisfied, which is true with the desired probability and that the δ_k's are chosen appropriately small as an increasing sequence with $\delta_2 < 1/2$. Then to prove the theorem we first need to choose η, as in the proof of Theorem 1, so that $\hat{t}(a)$ is within the range required for the inequalities of Lemma 1 to hold and then to note that for such $\hat{t}(a)$,

$$\left|\hat{K}^{(4)}(\hat{t}(a))\right|/\hat{K}^{(2)}(\hat{t}(a))^2 < C(1 + n^{-1/2+\delta_4}) \tag{3.6}$$

for $|a| < \eta n^{-\alpha}$. Then it remains only to show that

$$\mathrm{P}(E_n) > 1 - C' \exp(-cn^{\alpha}). \tag{3.7}$$

First, let E_n' and E_n'' be the sets on which (3.2) and (3.3) hold. Then from Lemma 1, $\mathrm{P}(E_n') > 1 - C' \exp(-cn^{\alpha})$ by the same methods as in the proof of Theorem 1.

To obtain such a result for E_n'' we proceed by a number of lemmas, the proofs of which we defer to the next subsection.

LEMMA 3. *Under the assumption* (3.5), *if* $\phi_B(\xi) = \mathrm{E}(\exp(i\xi X)\,|\,|X| < B)$ *then we can choose* B *sufficiently large such that* $|\phi_B(\xi)| < 1 - d/2$ *for all* $|\xi| > \gamma$.

LEMMA 4. *If* $N = \sum_{j=1}^n I(|X_j| < B)$ *then for sufficiently large* B *there is a constant* c *such that if* $E_n^* = \{N > n/2\}$ *then* $\mathrm{P}(E_n^*) > 1 - e^{-cn}$.

LEMMA 5. *If*

$$\hat{\phi}_B(\xi) = \frac{\sum_{j=1}^n \exp(i\xi X_j) I(|X_j| < B)}{\sum_{j=1}^n I(|X_j| < B)} \tag{3.8}$$

and if $E_n^{**} = \{|\hat{\phi}_B(\xi))| < 1 - d/4, \gamma < |\xi| < \Gamma n\}$, *then*

$$P(E_n^{**}) > 1 - e^{-cn}. \tag{3.9}$$

LEMMA 6. *Suppose*

$$\left|\hat{\phi}_B(\xi)\right| = \left|N^{-1} \sum_{j=1}^n \exp(i\xi X_j) I(|X_j| < B)\right| < 1 - d/4 \quad \text{for all } \gamma < |\xi| < \Gamma n.$$

Then given $\varepsilon > 0$ *there exists* $\delta > 0$ *such that for all real* x *and all* $\gamma < \xi < \Gamma n$,

$$|\xi X_j - x| > \varepsilon \tag{3.10}$$

for at least δN *of the* j*'s.*

To complete the proof, assume without loss of generality that $|X_j| < B$ for $j = 1, \ldots, N$ and that (3.10) holds for $j = 1, \ldots, \delta N$. Then

$$\begin{aligned}\left|\exp(\hat{K}(t + i\xi) - \hat{K}(t))\right| &= \left|\sum_{j=1}^{n} p_j(t)\exp(i\xi X_j)\right|,\\ &= 1 - \sum_{j=1}^{n} p_j(t)(1 - \cos(\xi X_j - y)),\end{aligned}$$

where

$$p_j(t) = \frac{\exp(tX_j)}{\sum_{j=1}^{n}\exp(tX_j)},$$

if we choose y such that

$$\tan y = \frac{\sum_{j=1}^{n} p_j(t)\sin(\xi X_j)}{\sum_{j=1}^{n} p_j(t)\cos(\xi X_j)} \quad \text{and} \quad \sum_{j=1}^{n} p_j(t)\cos(\xi X_j - y) \geq 0.$$

Now $|t| < A$, so for $j = 1, \ldots, N$,

$$\frac{b}{N} = \frac{e^{-AB}}{Ne^{AB}} \leq p_j(t).$$

So from Lemma 6,

$$\begin{aligned}\left|\exp(\hat{K}(t + i\xi) - \hat{K}(t))\right| &\leq 1 - \sum_{j=1}^{\delta N} p_j(t)(1 - \cos(\varepsilon))\\ &\leq 1 - b\delta(1 - \cos(\varepsilon)).\end{aligned}$$

Thus $E_n^* \cap E_n^{**} \subset E_n''$ and the theorem follows from Lemmas 4 and 5. □

3.2. *Proofs of lemmas.*

PROOF OF LEMMA 3. From (3.5) $|\phi(\xi)| < 1 - d$ for $d > 0$. So

$$\begin{aligned}|\phi_B(\xi) - \phi(\xi)| &= \left|\mathrm{E}(\exp(i\xi X)|\ |X| < B) - \mathrm{E}\exp(i\xi X)\right|\\ &= \left|\frac{\mathrm{E}\exp(i\xi X)I(|X| < B)}{\mathrm{E}\, I(|X| < B)} - \mathrm{E}\exp(i\xi X)\right|\\ &\leq 2(1 - P(|X| < B))\end{aligned}$$

and we can choose B to make this less than $d/2$. □

PROOF OF LEMMA 4. Take B such that $\mathrm{P}(|X| < B) \geq 3/4$, say. The result then follows immediately from Cramér's theorem applied to the sum of binary random variables. □

PROOF OF LEMMA 5. To see this note that

$$\begin{aligned}\left|\hat\phi_B(\xi)-\phi_B(\xi)\right| &= \left|\frac{\sum_{j=1}^n \exp(i\xi X_j)I(|X_j|<B)}{\sum_{j=1}^n I(|X_j|<B)} - \frac{\mathrm{E}\exp(i\xi X)I(|X|<B)}{\mathrm{E}\,I(|X|<B)}\right| \\ &\le \left|\frac{n^{-1}\sum_{j=1}^n \exp(i\xi X_j)I(|X_j|<B) - \mathrm{E}\exp(i\xi X)I(|X|<B)}{P(|X|<B)}\right| \\ &\quad + \left|n^{-1}\sum_{j=1}^n \exp(i\xi X_j)I(|X_j|<B)\right|\left|\frac{n}{N}-\frac{1}{P(|X|<B)}\right|.\end{aligned}$$

Let E_ξ be the set such that

$$\left|\hat\phi_B(\xi)-\phi_B(\xi)\right| < d/8, \tag{3.11}$$

then Cramér's theorem asserts that

$$\mathrm{P}(E_{\xi_0}) > 1-e^{-cn}. \tag{3.12}$$

Now

$$\left|\hat\phi_B(\xi)-\phi_B(\xi)-\hat\phi_B(\xi_0)-\phi_B(\xi_0)\right| \le |\xi-\xi_0|\left(n^{-1}\sum_{j=1}^n |X_j| + \mathrm{E}|X|\right). \tag{3.13}$$

Let $E_{1n}=\{n^{-1}\sum_{j=1}^n |X_j|<K\}$, for $K>\mathrm{E}|X_1|$, then since X_1 has a moment generating function for t in an open interval containing 0,

$$\mathrm{P}(E_{1n}) > 1-\exp(-c'n),$$

and in E_{1n} the right-hand side of (3.13) is bounded by $|\xi-\xi_0|2K \le d/8$ if $|\xi-\xi_0|\le d/(16K)$. Now choose $\xi_1,\dots,\xi_L$ such that they are $d/(32K)$ apart and centered in intervals $I_1,\dots,I_L$ of length $d/(32K)$ whose union covers (ε, Cn). Then $L=O(n)$ so

$$\mathrm{P}\big(E_{1n}\cap E_{\xi_1}\cap\cdots\cap E_{\xi_L}\big) > 1-\exp(-c''n).$$

Further

$$E_n^{**} \supset E_{1n}\cap E_{\xi_1}\cap\cdots\cap E_{\xi_L},$$

so

$$\mathrm{P}(E_n^{**}) > 1-e^{-cn}.$$

PROOF OF LEMMA 6. Otherwise, for some $\varepsilon>0$ and any $\delta>0$ we can find $(1-\delta)N$ of the j's with $|X_j|<B$ and $|\xi X_j - x - 2k\pi|\le\varepsilon$ for some integer k. So assuming without loss of generality that these are the first $(1-\delta)N$ of the

j's and writing $z = \exp(ix)$ we have

$$\begin{aligned}
\left|\hat{\phi}_B(\xi)\right| &= \left| N^{-1} \sum_{j=1}^{(1-\delta)N} \exp(i\xi X_j) + N^{-1} \sum_{j=(1-\delta)N+1}^{N} \exp(i\xi X_j) \right| \\
&\geq 1 - \delta - \left| N^{-1} \sum_{j=1}^{(1-\delta)N} \bigl(\exp(i\xi X_j) - z\bigr) \right| - \left| N^{-1} \sum_{j=(1-\delta)N+1}^{N} \exp(i\xi X_j) \right| \\
&\geq 1 - \delta - \varepsilon(1-\delta) - \delta.
\end{aligned}$$

But since δ is arbitrary, this contradicts the Cramér condition for $\hat{\phi}_B$.

REMARK. The proof of Lemma 6 follows that of a related result in Kokic (1988).

4. Bounds for empirical moments and generating functions. The present section proves Lemmas 1 and 2 and thereby provides a major part of the technical background for the proofs of the results in Section 2. Care must be taken to obtain the right rates of convergence in the bounds for the estimation error of moments of all orders, in particular in Lemma 7. Thus an application of Rosenthal's lemma [cf. Petrov (1995), Section 2.3] instead of Lemma 7 would lead to a term of order $(l+1)\alpha n^{\alpha} \log n$ instead of $l\alpha n^{\alpha} \log n$ in the exponent in (4.15), which would not suffice to prove the result.

In the multivariate case, the problem of generalizing the univariate results is to establish the uniformity in the argument t of the empirical moment generating function, $\hat{m}(t)$, since we cannot prove the simultaneous bound for the infinite sequence of mixed moments, analogous to Lemma 9 from the univariate case.

4.1. *Univariate case.* Throughout this section we let $X_1, \dots, X_n$ be independent and identically distributed real random variables with moments m_l for, at least, some integer values of l. The empirical moments are denoted

$$\hat{m}_l = \frac{1}{n} \sum_{i=1}^{n} X_i^l.$$

LEMMA 7. *Let l be a fixed integer, p a fixed even integer and assume that*

$$\left|\mathrm{E} X_i^k\right| \leq ck!\lambda^k, \qquad k = 1, \dots, lp, \tag{4.1}$$

for some constants $c > 0$ and $\lambda > 0$. Then

$$\mathrm{E}|\hat{m}_l - m_l|^p < \frac{1}{8}\eta_l^p pp! l^{(l-1)p}\bigl(n^{1-p} p^{(l-1)p} + (np)^{-p/2} 2^{(l-1)p}\bigr), \tag{4.2}$$

where $\eta_l = 2(c+1)e^{l+1}\lambda^l/l$.

For any real r with $1 \leq r \leq p$,

$$\mathrm{E}|\hat{m}_l - m_l|^r < e^{2l}\eta_l^r r l^{(l-1)r}\bigl(n^{1-r} r^{lr} + n^{-r/2} r^{r/2} 2^{(l-1)r}\bigr). \tag{4.3}$$

PROOF. The proof of (4.2) is divided into three steps: first the cumulants of $\hat{m}_l - m_l$ are bounded on the basis of (4.1); then the pth moment of $\hat{m}_l - m_l$ is bounded in terms of its cumulants and finally a certain sum entering this expression is bounded.

The inequalities

$$e^{-k}k^k < k! \leq k^k,$$

for any $k \in \mathbf{N}$, and

$$k!/m! \leq k^{k-m},$$

for $m < k$, will be used frequently in the proof.

STEP 1. Let χ_k denote the kth cumulant of $\hat{m}_l - m_l$. We now prove that from (4.1) we get

$$|\chi_k| < n^{1-k}(lk)!\tilde{\lambda}_l^k, \qquad 2 \leq k \leq p, \tag{4.4}$$

where $\tilde{\lambda}_l = (c+1)(e\lambda)^l/l$, while $\chi_1 = 0$ because the statistic is centered. If κ_k denotes the kth cumulant of X_i^l, we have

$$\chi_k = n^{1-k}\kappa_k. \tag{4.5}$$

The expression for κ_k in terms of the moments $\mu_k = \mathrm{E}(X_i^l)^k$ is

$$\kappa_k = k! \sum_{m=1}^{k} \frac{1}{m}(-1)^{m-1} \sum_{T(k,m)} \prod_{j=1}^{m} \frac{\mu_{a_j}}{a_j!},$$

where $T(k, m)$ is the set of sequences

$$T(k,m) = \Big\{(a_1, \dots, a_m) \in \mathbf{N}^m \colon \textstyle\sum a_j = k\Big\} \tag{4.6}$$

consisting of strictly positive integers. Insertion of the bounds (4.1) for $\mu_k = \mathrm{E}X_i^{lk}$ gives

$$\begin{aligned}
|\kappa_k| &\leq k! \sum_{m=1}^{k} \frac{1}{m} \sum_{T(k,m)} \prod_{j=1}^{m} \left(\frac{c\lambda^{la_j}(la_j)!}{a_j!}\right) \\
&\leq k!\lambda^{lk} \sum_{m=1}^{k} \frac{c^m}{m} \sum_{T(k,m)} \prod_{j=1}^{m} (la_j)^{(l-1)a_j} \\
&\leq k!\lambda^{lk} l^{(l-1)k} k^{(l-1)k} \sum_{m=1}^{k} \frac{c^m}{m} |T(k,m)|,
\end{aligned}$$

where $|T(k,m)| = \binom{k-1}{m-1}$ denotes the number of sequences in $T(k,m)$, and we have used $(la_j)!/a_j! \le (la_j)^{la_j - a_j}$ to obtain the second inequality. Thus,

$$\begin{aligned} |\kappa_k| &\le k!\lambda^{lk} l^{(l-1)k} k^{(l-1)k} c(c+1)^{k-1} \\ &< \left(\lambda^l (c+1)/l\right)^k (lk)^{lk} \\ &< (lk)! e^{lk} \left(\lambda^l (c+1)/l\right)^k, \end{aligned}$$

which gives (4.4) when substituted into (4.5).

STEP 2. Using the bounds just obtained for the cumulants, $\{\chi_k\}$, we continue to prove that

$$\mathrm{E}\left|\hat{m}_l - m_l\right|^p < p!\tilde{\lambda}_l^p l^{(l-1)p} n^{-p} 2^{p-2} G(n,p,m), \tag{4.7}$$

where

$$G(n,p,m) = \sum_{m=1}^{p/2} \frac{n^m}{m!} 2^{2(l-1)(m-1)} (p-2m+2)^{(l-1)(p-2m+2)}. \tag{4.8}$$

The proof of (4.7) starts from the expression

$$\mathrm{E}(\hat{m}_l - m_l)^p = p! \sum_{m=1}^{p} \frac{1}{m!} \sum_{T(p,m)} \prod_{j=1}^{m} \frac{\chi_{a_j}}{a_j!}$$

for the pth moment in terms of cumulants. Since $\chi_1 = 0$ we shall use the notation $\sum^*$ for the sum over the set of sequences

$$T^*(p,m) = \left\{(a_1, \dots, a_m) \in T(p,m)\colon a_j \ge 2, 1 \le j \le m\right\}, \tag{4.9}$$

which must have $m \le p/2$ to be nonempty. Insertion of the bounds (4.4) now gives

$$\begin{aligned} \mathrm{E}(\hat{m}_l - m_l)^p &= p! \sum_{m=1}^{p/2} \frac{1}{m!} \sum{}^* \prod_{j=1}^{m} \frac{n^{1-a_j}(la_j)!\tilde{\lambda}_l^{a_j}}{a_j!} \\ &\le p!\tilde{\lambda}_l^p l^{(l-1)p} \sum_{m=1}^{p/2} \frac{n^{m-p}}{m!} \sum{}^* \prod_{j=1}^{m} a_j^{(l-1)a_j}. \end{aligned} \tag{4.10}$$

A sequence $(a_1, \dots, a_m) \in T^*(p,m)$ may be regarded as an m-point distribution with mean $\sum a_j/m = p/m$. The distribution is concentrated on the interval $(2, p-2m+2)$ since $a_j \ge 2$ for all j. The maximum of $\mathrm{E}(Y \log Y)$ over all distributions of Y on this interval with mean p/m puts all mass at the endpoints, since the function $y \mapsto y \log y$ is convex. The maximum is obtained by the distribution given by

$$a_1 = p - 2m + 2, \qquad a_2 = \cdots = a_m = 2.$$

Since $a_j^{(l-1)a_j} = \exp\{(l-1)a_j \log a_j\}$, this leads to the bound

$$\prod a_j^{(l-1)a_j} = \exp\left((l-1)\sum a_j \log a_j\right) \le 2^{2(l-1)(m-1)} (p-2m+2)^{(l-1)(p-2m+2)}.$$

Substitution into (4.10) together with the combinatorial observation that the number of elements in $T^*(p, m)$ is $\binom{p-m-1}{m-1} < 2^{p-m-1} \le 2^{p-2}$ leads to inequality (4.7).

STEP 3. For the sum $G(n, p, m)$ from (4.8) we now prove that

$$G(n, p, m) < \tfrac{1}{2} p e^p \{ n p^{(l-1)p} + (n/p)^{p/2} 2^{(l-1)p} \} \tag{4.11}$$

from which the result (4.2) follows immediately. To prove (4.11), observe first that $1/m! < \exp(m - m \log m)$. Let $u = m/p$ such that

$$m \log m = m \log p + pu \log u \ge m \log p - pe^{-1},$$

because the convex function $u \mapsto u \log u$ has a minimum equal to $-e^{-1}$ at $u = e^{-1}$ for $u > 0$. Thus,

$$G(n, p, m) < \sum_{m=1}^{p/2} e^{h(m)}, \tag{4.12}$$

where

$$\begin{aligned} h(m) = m \log n + m - m \log p + pe^{-1} + 2(l-1)(m-1) \log 2 \\ + (l-1)(p - 2m + 2) \log(p - 2m + 2) \end{aligned}$$

is convex. Hence $\exp\{h(m)\}$ is convex and positive, so

$$\begin{aligned} \exp(h(m)) &< \exp(h(1)) + \exp(h(p/2)) \\ &= \exp(pe^{-1}) \{ nep^{-1} p^{(l-1)p} + (ne/p)^{p/2} 2^{(l-1)p} \} \end{aligned}$$

for $1 \le m \le p/2$. Multiplication by the number of terms, $p/2$, now easily gives the bound (4.11) for $G(n, p, m)$. This completes the proof of inequality (4.2) in Lemma 7.

To obtain inequality (4.3) we use

$$\mathrm{E}|\hat{m}_l - m_l|^r \le \left(\mathrm{E}|\hat{m}_l - m_l|^p\right)^{r/p}, \qquad r \le p,$$

with p taken as an even integer satisfying $r \le p < r + 2$. This leads to the desired inequality after use of the estimations

$$(A + B)^{r/p} \le A^{r/p} + B^{r/p}, \qquad A, B > 0$$

and

$$p!^{r/p} \le p^r < r^r e^2,$$

and $p^{-r/2} \le r^{-r/2}$ together with some manipulations with constants. This completes the proof of Lemma 7. □

REMARK. The result (4.2) is actually proved under the single condition that the statistic has cumulants satisfying the inequalities (4.4) without further use of the structure of the statistic as a sum of centered powers.

LEMMA 8. *Assume that X_i has finite moment generating function in a neighborhood of the origin. For any $l \in \mathbf{N}$, $C > 0$ and $c > 0$, and any constants $\alpha > 0$ and $\delta_l > 0$ satisfying the inequalities*

$$\tfrac{1}{2} + \delta_l - l\alpha > 0, \tag{4.13}$$

$$\delta_l - \tfrac{1}{2}\alpha > 0, \tag{4.14}$$

there exists a $c_1 > 0$ and a $C' > 0$, both independent of l, such that

$$\begin{aligned}
&\mathrm{P}\big(|\hat{m}_l - m_l| \le Cn^{-1/2+\delta_l}\big)\\
&\quad > 1 - \exp\big(c_1 l n^\alpha + (l-1)n^\alpha \log l\big) \qquad (4.15)\\
&\qquad \times \big(\exp\big(-(1/2 + \delta_l - l\alpha)n^\alpha \log n\big) + \exp\big(-(\delta_l - \alpha/2)n^\alpha \log n\big)\big)\\
&\quad > 1 - C' \exp(-cn^\alpha). \qquad (4.16)
\end{aligned}$$

PROOF. Since X_i has finite moment generating function in a neighborhood of the origin, condition (4.1) is fulfilled for all $k \in \mathbf{N}$ for some c and λ. By Chebyshev's inequality and Lemma 7 with $r = n^\alpha$,

$$\begin{aligned}
&\mathrm{P}\big(|\hat{m}_l - m_l| \le Cn^{-1/2+\delta_l}\big)\\
&\quad \ge 1 - C^{-r} n^{r/2-\delta_l r} \mathrm{E}|\hat{m}_l - m_l|^r\\
&\quad > 1 - C^{-r} n^{r/2-\delta_l r} \exp(c_1 l r) l^{(l-1)r}\big(n^{1-r} r^{lr} + n^{-r/2} r^{r/2}\big)\\
&\quad \ge 1 - \exp(c_1 l r + n^\alpha (l-1)\log l)\\
&\qquad \times \big(\exp(-(1/2 + \delta_l - l\alpha)n^\alpha \log n) + \exp(-(\delta_l - \alpha/2)n^\alpha \log n)\big)
\end{aligned}$$

for suitable $c_1 > 0$. This proves the first inequality of the lemma; the second follows easily. □

Lemma 2 trivially extends to a simultaneous bound for moments of order $l = 1, \ldots, L$ by summing up the probabilities of failure for each l. The following lemma shows that a simultaneous bound holds for all moments, that is, with $L = \infty$.

LEMMA 9. *Assume that X_i has finite moment generating function in a neighborhood of the origin and let $c > 0$, $C > 0$ and $L \in \mathbf{N}$ be arbitrary but fixed. Let $\alpha, \delta_1, \delta_2, \ldots,$ be any positive constants satisfying the conditions* (4.13) *and* (4.14) *for $l = 1, \ldots, L$, and*

$$\delta_l = \delta_{l,n} = \delta_L + (l - L)\big\{\alpha + (\beta + \log l)/\log n\big\} \tag{4.17}$$

for $l = L+1, L+2, \ldots$ and some sufficiently large constant β. Then there exists a $C' > 0$ such that

$$\mathrm{P}\big(|\hat{m}_l - m_l| \le Cn^{-1/2+\delta_l}; l = 1, 2, \ldots\big) > 1 - C' \exp(-cn^\alpha). \tag{4.18}$$

PROOF. Let P_L denote the probability on the left-hand side in (4.18), except that we consider only $l = L, L+1, \dots,$ since any finite number of terms may be disregarded. From Lemma 8 we have the bound

$$P_L > 1 - \sum_{l=L}^{\infty} \exp(c_1 l n^{\alpha} + (l-1) n^{\alpha} \log l) \times \bigl(\exp(-(1/2 + \delta_l - l\alpha) n^{\alpha} \log n) + \exp(-(\delta_l - \alpha/2) n^{\alpha} \log n)\bigr).$$

Let $r_L = \min\{1/2 + \delta_L - L\alpha, \delta_L - \alpha/2\}$, which is positive. Then,

$$P_L > 1 - 2 \sum_{l=L}^{\infty} \exp(l(c_1 - \beta) n^{\alpha} + (\log l)(L-1) n^{\alpha} + L\beta n^{\alpha} - r_L n^{\alpha} \log n).$$

If $\beta > c_1$, the exponential factor $\exp\{-l(\beta - c_1) n^{\alpha}\}$ will dominate the tail behavior of the sum, which will then behave like its leading term so that

$$P_L > 1 - C'' \exp(L(c_1 - \beta) n^{\alpha} + (\log L)(L-1) n^{\alpha} + L\beta n^{\alpha} - r_L n^{\alpha} \log n)$$

for some $C'' > 0$. As $n \to \infty$ the last term in the exponent becomes dominant, thus proving the result. □

The rate constant δ_l increases as αl with l in Lemma 9. Thus, for higher order moments, Lemma 9 does not even prove convergence of the empirical moments with the given probability. However, the result is essentially the same as the one established in Lemma 8, which was not a simultaneous bound, and it is strong enough to obtain uniform bounds for the moment generating function and its derivatives as shown in the following lemma.

LEMMA 10. *Under the conditions of Lemma* 9, *for any* $C' > 0$ *and* $c > 0$ *there exist positive constants* C'', C_k *and* η, *such that*

$$\mathrm{P}\bigl(\bigl|\hat{m}^{(k)}(t) - m^{(k)}(t)\bigr| < C_k n^{-1/2 + d_k}; k = 0, 1, \dots; |t| < \eta n^{-\alpha}\bigr) > 1 - C'' \exp(-c n^{\alpha}), \tag{4.19}$$

where $m^{(k)}(t)$ *denotes the kth derivative of the moment generating function*, $C_k = C' k! \exp(c' k)$, *and*

$$d_k = \begin{cases} \max\{\delta_k, \delta_{k+1} - \alpha, \dots, \delta_L - (L-k)\alpha\}, & \text{for } k \le L, \\ \delta_k, & \text{for } k > L, \end{cases} \tag{4.20}$$

where δ_0 *should be omitted on the right in the definition of* d_0.

For $k = 0$ *the bound for* $|\hat{m}(t) - m(t)|$ *may be sharpened to* $C' n^{-1/2 + d_1} |t|$.

REMARK. Notice that the d_k's satisfy the same conditions (4.13) and (4.14) as the δ_k's.

PROOF. A Taylor series expansion around $t = 0$ shows that with the same probability as in Lemma 9 we have for any $C > 0$,

$$\left|\hat{m}^{(k)}(t) - m^{(k)}(t)\right| \leq C \sum_{l=k}^{\infty} \frac{1}{(l-k)!} n^{-1/2+\delta_l} |t|^{l-k},$$

where the convergence of the Taylor series is guaranteed within the range of $|t|$, which gives geometric convergence, because the moment generating function is known to be analytic. Assume first that $k \leq L$. Then,

$$\delta_l - d_k \leq \delta_l - (\delta_l - (l-k)\alpha) = (l-k)\alpha,$$

for $l \leq L$, so that we may bound the first part of the sum by

$$\sum_{l=k}^{L} \leq Cn^{-1/2+d_k} \sum_{j=0}^{L-k} \frac{1}{j!} \{n^\alpha |t|\}^j < n^{-1/2+d_k} e^\eta$$

for $|t| \leq \eta n^{-\alpha}$. For the tail part of the sum from $l = L + 1$ we get the bound

$$\begin{aligned}
\sum_{l=L+1}^{\infty} &\leq Cn^{-1/2+\delta_L} \sum_{l=L+1}^{\infty} \frac{1}{(l-k)!} n^{(l-L)\alpha + (l-L)(\beta + \log l)/\log n} |t|^{l-k} \\
&= Cn^{-1/2+d_L} |t|^{L-k} \sum_{l=L+1}^{\infty} \frac{1}{(l-k)!} \left(n^\alpha e^\beta |t|\right)^{l-L} l^{l-L} \\
&< Cn^{-1/2+d_L} |t|^{L-k} e^L \sum_{l=L+1}^{\infty} \left(n^\alpha e^\beta |t|\right)^{l-L},
\end{aligned}$$

where we have used the inequality $l^{l-L} < (l-L)! e^L$ to obtain the last inequality. Taking C sufficiently small, this proves the result for $k \leq L$. For the special case $k = 0$, the term $l = 0$ disappears from the sum because $\hat{m}(0) = m(0)$, thus leading to trivial modifications.

For $k \geq L + 1$ we similarly get the bound

$$Cn^{-1/2+d_L+(k-L)\alpha} \exp((k-L)\beta) \sum_{l=k}^{\infty} \frac{1}{(l-k)!} \left(n^\alpha e^\beta |t|\right)^{l-k} l^{l-L}.$$

Using again the inequality $l^{l-L} < (l-L)! e^L$ together with

$$\sum_{l=k}^{\infty} \frac{(l-L)!}{(l-k)!} a^{l-k} = (k-L)!(1-a)^{L-k}$$

for $|a| < 1$, the result of the lemma follows. □

PROOF OF LEMMA 1. It suffices to prove the result for any fixed k. To do this we use Lemma 10 with

$$\delta_l = \max\left\{\tfrac{1}{2}\alpha, l\alpha - \tfrac{1}{2}\right\} + \varepsilon,$$

where $\varepsilon > 0$ should be chosen sufficiently small to guarantee that δ_k does not exceed the δ_k originally given in the statement of Lemma 1 and in any case with $\varepsilon < (1-\alpha)/2$.

In the sequel we assume that the bounds for $|\hat{m}_l(t) - m_l(t)|$ obtained in Lemma 10 hold; this is true with the desired probability. Note that with the present choice of δ_k's we get $d_k = \delta_k$ in Lemma 10.

Consider first the case $k = 0$. For $|t| < \eta' n^{-\alpha}$,

$$\big|\hat{K}(t) - K(t)\big| = \big|\log \hat{m}(t) - \log m(t)\big| \le C''\big|\hat{m}(t) - m(t)\big|,$$

for some $C'' > 0$, because $m(t)$ converges uniformly to 1 in the range considered, and $|\hat{m}(t) - m(t)|$ converges uniformly to 0, because $\delta_1 < 1/2$. Choosing C' in Lemma 10 sufficiently small this proves the case $k = 0$.

For $k \ge 1$ notice that $K^{(k)}(t)$ is a linear combination of terms of the form

$$\frac{m^{(k_1)}(t)\cdots m^{(k_s)}(t)}{m(t)^s},$$

with $1 \le s \le k$, $k_j \ge 1$ and $\sum k_j = k$. Any difference of the form

$$\big|\hat{m}^{(k_1)}(t)\cdots\hat{m}^{(k_s)}(t) - m^{(k_1)}(t)\cdots m^{(k_s)}(t)\big|$$

may be bounded by a linear combination of terms of the form

$$\begin{aligned} &\big|\hat{m}^{(k_1)}(t)\cdots\hat{m}^{(k_{j-1})}(t)\big|\,\big|\hat{m}^{(k_j)}(t) - m^{(k_j)}(t)\big|\,\big|m^{(k_{j+1})}(t)\cdots m^{(k_s)}(t)\big| \\ &\quad \le C'' n^{\delta_{k_j} - 1/2 + \sum_{i\ne j}\max\{0, \delta_{k_i} - 1/2\}}, \end{aligned} \tag{4.21}$$

where any $C'' > 0$ will be valid with probability $1 - C'\exp(-cn^\alpha)$ for some C'. We continue by showing that for $\sum_{i=1}^{s'} k_i \le k'$ we have

$$\sum_{i=1}^{s'}\big(\delta_{k_i} - \tfrac{1}{2}\big) \le \delta_{k'} - \tfrac{1}{2}. \tag{4.22}$$

Observe that δ_l's for which $\delta_l = \frac{1}{2}\alpha + \varepsilon$ satisfy $\delta_l - \frac{1}{2} < 0$ and may hence be disregarded on the left in (4.22). Thus we may assume that $\delta_l = l\alpha - \frac{1}{2} + \varepsilon$ for $l = k_1, \dots, k_{s'}$ in (4.22). This gives

$$\begin{aligned} \sum_{i=1}^{s'}\big(\delta_{k_i} - \tfrac{1}{2}\big) &= \sum_{i=1}^{s'}(k_i\alpha - 1 + \varepsilon) \\ &\le (k'\alpha - 1 + \varepsilon) - (s' - 1)(1 - \varepsilon) \\ &\le \delta_{k'} - \tfrac{1}{2}. \end{aligned}$$

Thus, with $k' = k - k_j$ in (4.21), we get the bound

$$C'' n^{\delta_{k_j} - 1/2 + \delta_{k'} - 1/2} < C'' n^{\delta_k - 1/2}$$

by the same argument as above.

We further need to bound differences like

$$\frac{m^{(k_1)}(t)\cdots m^{(k_s)}(t)}{\hat{m}(t)^s} - \frac{m^{(k_1)}(t)\cdots m^{(k_s)}(t)}{m(t)^s}, \tag{4.23}$$

but the factors $m^{(k_j)}(t)$ are uniformly bounded, $m(t)$ converges uniformly to 1 and

$$\left|\hat{m}(t) - m(t)\right| \le n^{\delta_1 - 1/2 - \alpha} < n^{\delta_k - 1/2},$$

so the difference (4.23) will also be of this order.

4.2. *Multivariate case.*

PROOF OF LEMMA 2. The proof relies heavily on the results obtained for the univariate case, but their extension is not trivial. If $v \in \mathbf{R}^d$ is any fixed vector, the random variables $v^T X_1, \dots, v^T X_n$ satisfy the conditions of Lemmas 7–10, so, in particular, Lemma 10 ensures that

$$\mathrm{P}\big(|\hat{m}(sv) - m(sv)| < Cn^{-1/2 + d_1 - \alpha}; |s| < \eta n^{-\alpha}\big) > 1 - C' \exp(-cn^{\alpha}),$$

where $s \in \mathbf{R}$, because $m(sv) = \mathrm{E}\exp(sv^T X_1)$ is the moment generating function of $v^T X_1$. It is also clear that we get a bound of this type for any finite number of vectors v; we may even include a number of vectors growing as any power of n, because the exponential rate in n^α will still dominate. The problem is, however, to obtain the result uniformly for all v with $\|v\| = 1$, say, and similarly for the derivatives of $\hat{m}(t) - m(t)$.

We continue by showing that for any finite number of partial derivatives we have similarly the result

$$\begin{aligned} &\mathrm{P}\big(|\hat{m}_{i_1\cdots i_k}(sv) - m_{i_1\cdots i_k}(sv)| < Cn^{-1/2 + d_k}; 1 \le k \le L; |s| < \eta n^{-\alpha}\big) \\ &\qquad > 1 - C' \exp(-cn^{\alpha}), \end{aligned} \tag{4.24}$$

where $L \in \mathbf{N}$, v is fixed with $\|v\| = 1$, and it is understood that the i_j's may denote any of the coordinates. Note that Lemma 10 does not give this bound, because it only involves derivatives with respect to s, so other methods must be used.

For notational convenience we take $k = 2$, $i_1 = 1$ and $i_2 = 2$ as an example. Let $t = sv$, and let $m_{12}^{(l)}(0)$ denote the lth derivative of $m_{12}(sv)$ with respect to s at $s = 0$. Then

$$\begin{aligned} |\hat{m}_{12}(t) - m_{12}(t)| &\le \sum_{l=1}^{L} \frac{s^l}{l!}\big|\hat{m}_{12}^{(l)}(0) - m_{12}^{(l)}(0)\big| \\ &\quad + \sum_{l=L+1}^{\infty} \frac{s^l}{l!}\big|\hat{m}_{12}^{(l)}(0) - m_{12}^{(l)}(0)\big|, \end{aligned} \tag{4.25}$$

The first sum only involves finitely many empirical mixed moments of X_{i1}, X_{i2} and $v^T X_i$, since $m_{12}^{(l)}(0) = \mathrm{E}\{X_{i1} X_{i2} (v^T X_i)^l\}$ is a mixed moment of order

$l+2$. The polarization identity [see Federer (1969), Section 1.9.3] may be used to show that any such moment may be written as a linear combination of moments of order $l+2$ of all the univariate variables of the form $u^T X_i$ with

$$u^T X_i = \pm X_{i1} \pm X_{i2} \pm v^T X_i \pm \cdots \pm v^T X_i,$$

with $l+2$ vectors on the right-hand side; see also Skovgaard [(1990), Section 1.4 and equation (1.6)]. The number of vectors u arising in this way is 2^{l+2}, their lengths do not exceed $l+2$, and the coefficients are all $\{(l+2)!2^{l+2}\}^{-1}$ in absolute value. Hence, the estimation error for the mixed moments may be bounded as in Lemma 8 with their order $l+2$ replacing l in the lemma, and the first sum in (4.25) may be bounded as in the proof of Lemma 10 with a leading term of order $n^{-1/2+\delta_2}$ or $n^{-1/2+\delta_k}$ for the more general case with a partial derivative of order k.

For the second sum in (4.25) we use first the inequality

$$\begin{aligned}\left|\hat{m}_{12}^{(l)}(0) - m_{12}^{(l)}(0)\right| &\le \left|\hat{m}_{12}^{(l)}(0)\right| + \left|m_{12}^{(l)}(0)\right| \\ &= \left|\hat{\mathrm{E}}\{X_{i1} X_{i2} (v^T X_i)^l\}\right| + \left|m_{12}^{(l)}(0)\right|.\end{aligned}$$

The second term on the right is bounded, so if L is chosen large enough it will be outweighed by s^l. For the first term we use Hölder's inequality,

$$\begin{aligned}&\hat{\mathrm{E}}(|X_{i1}|\,|X_{i2}|\,|v^T X_i|^l) \\ &\quad\le \left(\hat{\mathrm{E}}|X_{i1}|^{l+2}\right)^{1/(l+2)} \left(\hat{\mathrm{E}}|X_{i2}|^{l+2}\right)^{1/(l+2)} \left(\hat{\mathrm{E}}|v^T X_i|^{l+2}\right)^{l/(l+2)} \\ &\quad\le \left(\hat{\mathrm{E}} X_{i1}^r\right)^{1/r} \left(\hat{\mathrm{E}} X_{i2}^r\right)^{1/r} \left(\hat{\mathrm{E}}(v^T X_i)^r\right)^{l/r},\end{aligned}$$

where r is an even integer with $r-2 < l+2 \le r$. Each of the three moments of the last expression is of order r and hence bounded in order of magnitude by $\max\{1, n^{-1/2+\delta_r}\}$, using Lemma 8. Notice that we only use these bounds for the finite number of random variables consisting of the coordinates and $v^T X_i$, so a simultaneous bound applies with the desired probability. The second sum in (4.25) may now be bounded as in the proof of Lemma 9, because the bound for $|\hat{m}_{12}^{(l)}(0) - m_{12}^{(l)}(0)|$ is of order

$$\left(n^{-1/2+\delta_r}\right)^{1/r} \left(n^{-1/2+\delta_r}\right)^{1/r} \left(n^{-1/2+\delta_r}\right)^{l/r} = \left(n^{-1/2+\delta_r}\right)^{(l+2)/r},$$

or of order 1 for small values of r. Thus, the factor $s^l/l!$ ensures that the sum is finite and tends sufficiently rapidly to zero to establish (4.24) for the special case with $k=2$. Other cases are proved in exactly the same way.

We have now proved that (4.24) holds for any fixed unit vector v and hence also simultaneously for any n^m preselected unit vectors, for any m. To show that it then also holds with sv replaced by any $t \in \mathbf{R}^d$ within a range of the form $\|t\| < \eta n^{-\alpha}$ we first note that the fixed vectors may be chosen at equidistant angles such that t will be within a distance of order n^{-m} from a vector sv, where v is one of the preselected vectors. Let $h = t - sv$ and use again the second derivative with respect to the two first coordinates as example. Then

$$
\begin{aligned}
|\hat{m}_{12}(t)-\hat{m}_{12}(sv)| &= \left|\frac{1}{n}\sum_{i=1}^{n} X_{i1}X_{i2}\{\exp(t^TX_i)-\exp(sv^TX_i)\}\right| \\
&= \left|\frac{1}{n}\sum_{i=1}^{n} X_{i1}X_{i2}\exp(sv^TX_i)\{\exp(h^TX_i)-1\}\right| \\
&\le \frac{1}{n}\sum_{i=1}^{n} |X_{i1}X_{i2}|\exp(sv^TX_i)|h^TX_i|\exp(|h^TX_i|),
\end{aligned}
$$

using the inequality $|e^y-1|\le |y|\,e^{|y|}$. By the Cauchy–Schwarz inequality the bound becomes

$$
\begin{aligned}
&\left(\frac{1}{n}\sum_{i=1}^{n}(X_{i1}X_{i2})^2(h^TX_i)^2\exp(2sv^TX_i)\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n}\exp(2\,|h^TX_i|)\right)^{1/2} \\
&\qquad \le \left(\frac{1}{n}\sum_{i=1}^{n}(X_{i1}X_{i2})^2(h^TX_i)^2\exp(2sv^TX_i)\right)^{1/2}\{\hat{m}(2h)+\hat{m}(-2h)\}^{1/2}.
\end{aligned}
\tag{4.26}
$$

For the first factor we write (h^TX_i) in coordinates and get the bound

$$
\left(\max_j |h_j|\right)\left(\sum_{i=1}^{d}\sum_{j=1}^{d}\hat{m}_{1122ij}(2sv)\right)^{1/2} \le C\left(\max_j |h_j|\right)(d^2n^{-1/2+\delta_6})^{1/2},
$$

where h_j is the jth coordinate of h and we have used that v is one of the preselected directions. Since $\|h\|<n^{-m}$, where m may be chosen arbitrarily large, the bound is of the order of any negative power of n. In (4.26) each of $\hat{m}(2h)$ and $\hat{m}(-2h)$ may be bounded by use of a convexity argument; the vector $2h$ may be written as a convex combination of d vectors proportional to the coordinate vectors, with coefficient $(\sum|h_j|)\,\mathrm{sign}(h_i)$ for the ith coordinate vector. Each of these vectors has length bounded by a constant times $\|h\|$, so Jensen's inequality applied to the convex function $\log m(t)$, ensures that $m(2h)$ and $m(-2h)$ are bounded.

Putting it all together, the critical term on the right side of the inequality

$$
\begin{aligned}
&|\hat{m}_{12}(t)-m_{12}(t)| \\
&\qquad \le |\hat{m}_{12}(t)-\hat{m}_{12}(sv)|+|\hat{m}_{12}(sv)-m_{12}(sv)|+|m_{12}(t)-m_{12}(sv)|
\end{aligned}
$$

is the middle one, which is bounded by $Cn^{-1/2+\delta_2}$ above. Other partial derivatives being similar, this completes the proof of the analogue of Lemma 10 for the multivariate case.

The transition of the estimation error for the moment generating function and its derivatives to that of the cumulant generating function and its derivatives follows the same lines as in the univariate case, for which this was the essential part of the proof of Lemma 1. We omit the details which involve notationally complicated relations between multivariate cumulants and moments; see, for example, McCullagh [(1987), Section 2.3]. More terms occur in the expressions but only the orders matter, and they are given by the orders of the moments involved, exactly as in the univariate case. □

Acknowledgment. We thank the referees, whose careful reading and helpful suggestions have allowed us to reformulate the results in the present improved form.

REFERENCES

ALBERS, W., BICKEL, P. J. and VAN ZWET, W. R. (1976). Asymptotic expansions for the power of distribution free tests in the one sample problem. *Ann. Statist.* **4** 108–156.

BARNDORFF-NIELSEN, O. (1986). Inference on full or partial parameters based on the standardized signed likelihood ratio. *Biometrika* **73** 307–322.

BICKEL, P. J. and VAN ZWET, W. R. (1978). Asymptotic expansions for the power of distribution free tests in the two sample problem. *Ann. Statist.* **6** 937–1004.

DANIELS, H. E. and YOUNG, G. A. (1991). Saddlepoint approximation for the studentized mean, with an application to the bootstrap. *Biometrika* **78** 169–179.

DAVISON, A. C. and HINKLEY, D. V. (1988). Saddlepoint approximations in resampling methods. *Biometrika* **75** 417–431.

DICICCIO, T. J., FIELD, C. A. and FRASER, D. A. S. (1990). Approximations of marginal tail probabilities and inference for scale parameters. *Biometrika* **77** 77–95.

DICICCIO, T. J., MARTIN, M. A. and YOUNG, G. A. (1992). Fast and accurate approximate double bootstrap confidence intervals. *Biometrika* **79** 285–295.

FEDERER, H. (1969). *Geometric Measure Theory*. Springer, Berlin.

FEUERVERGER, A. (1989). On the empirical saddlepoint approximation. *Biometrika* **76** 457–464.

HALL, P. (1990). On the relative performance of the bootstrap and Edgeworth approximations of a distribution function. *J. Multivariate Anal.* **35** 108–129.

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

JENSEN, J. L. (1992). The modified signed log likelihood statistic and saddlepoint approximations. *Biometrika* **79** 693–703.

JING, B. Y., FEUERVERGER, A. and ROBINSON, J. (1994). On the bootstrap saddlepoint approximations. *Biometrika* **81** 211–215.

JING, B. Y. and ROBINSON, J. (1994). Saddlepoint approximations for marginal and conditional probabilities of transformed variables. *Ann. Statist.* **22** 1115–1132.

KOKIC, P. N. (1988). Limit theorems for U-statistics in finite populations. Ph.D. thesis, Univ. Sydney.

MCCULLAGH, P. (1987). *Tensor methods in statistics*. Chapman and Hall, London.

PETROV, V. V. (1995). *Limit Theorems of Probability Theory.* **4** Clarendon Press, Oxford.

ROBINSON, J., HÖGLUND, T., HOLST, L. and QUINE, M. P. (1990). On approximating probabilities for small and large deviations in $\Re^d$. *Ann. Probab.* **18** 727–753.

SKOVGAARD, I. M. (1990). *Analytic Statistical Models*. IMS, Hayward, CA.

SCHOOL OF MATHEMATICS
AND STATISTICS
UNIVERSITY OF SYDNEY
NSW 2006
AUSTRALIA
E-MAIL: johnr@maths.oz.su.au

DEPARTMENT OF MATHEMATICS
AND PHYSICS
ROYAL VETERINARY
AND AGRICULTURAL UNIVERSITY
THORVALDSENSVEJ 40
DK-1871 FREDERIKSBERG C
DENMARK
E-MAIL: ims@kvl.dk