The Annals of Probability
1983, Vol. 11, No. 2, 346–354

ON THE ORDER OF MAGNITUDE OF CUMULANTS OF VON MISES FUNCTIONALS AND RELATED STATISTICS

BY R. N. BHATTACHARYA[1] AND M. L. PURI[2]

Indiana University

It is shown that under appropriate conditions the sth cumulant of a von Mises statistic or a U (or V) statistic is $O(n^{-s+1})$, $s \geq 2$, as the sample size n goes to infinity. A possible route toward the derivation of an asymptotic expansion of the characteristic function is indicated.

1. Introduction. The Edgeworth expansion of the characteristic function of a normalized sum of n independent and identically distributed (i.i.d) random variables derives from the order of magnitude $O(n^{-(s-2)/2})$ of the sth *cumulant* ($s \geq 2$) (See, e.g., Bhattacharya, 1977). For statistics which may be expressed as or approximated by polynomials in several average sample characteristics (e.g., (i) polynomials in sample moments and (ii) maximum likelihood estimators in the regular case), the validity of the so-called "formal Edgeworth expansion" depends crucially on the above order of magnitude of the sth cumulant ($s \geq 2$) of the normalized statistic (see Bhattacharya and Ghosh, 1978). In this note it is shown that cumulants of normalized U-*statistics* and *von Mises functionals* have the above order of magnitude, if certain conditions are satisfied. For general background on these statistics we refer to von Mises (1947) and Serfling (1980). Assuming the validity of (a) the above order of magnitude of the cumulants and (b) the Edgeworth expansion of the distribution function of a von Mises functional, Withers (1980) has given an algorithm for computing the coefficients in the asymptotic expansion. Some of the moment computations in Section 2 are similar to those in Withers (loc. cit). In Section 3 a new method of derivation of Cramér-Edgeworth expansions of characteristic functions of a class of statistics is provided.

2. Moments and cumulants. Let χ be a separable metric space (e.g., a subset of R^d), $\mathscr{B}_\chi$ its Borel sigma field, and P a given probability measure on $\mathscr{B}_\chi$, whose *support* is S. Let $\mathscr{P}_f$ denote the set of all probability measures on $\mathscr{B}_\chi \cap S$ having *finite supports*. Endow $\mathscr{P}_f \cup \{P\}$ with the weak-star topology. Consider for each n the product space $(\chi^n, \mathscr{B}_{\chi^n})$, and let $X_1, \cdots, X_n$ be the n coordinate random variables. Let $G^{\otimes n} = G \times G \times \cdot\cdot \times G$ denote the product probability measure on $\mathscr{B}_{\chi^n}$, where G is a probability measure on $\mathscr{B}_\chi$. Under $G^{\otimes n}$ the random variables $X_1, \cdots, X_n$ are i.i.d. with common distribution G. We shall write E_G to denote expectation under $G^{\otimes n}$. Denote the *empirical distribution* of the "observations" $X_1, \cdots, X_n$ by F_n, i.e., $F_n = n^{-1}\sum_{i=1}^n \delta_{X_i}$, where δ_x is the *Dirac measure* with point mass at x.

Let $h(x_1, x_2, \cdots, x_r)$ be a real-valued, Borel measurable, symmetric function on χ^r, for some $r \geq 2$. Define the V-*statistic* (with *kernel* h)

$$V_n = n^{-r} \sum_{i_1=1}^n \cdots \sum_{i_r=1}^n h(X_{i_1}, X_{i_2}, \cdots, X_{i_r}), \tag{2.1}$$

and the U-*statistic* (with *kernel* h)

$$U_n = \binom{n}{r}^{-1} \sum h(X_{i_1}, X_{i_2}, \cdots, X_{i_r}) \tag{2.2}$$

where the summation is over $1 \leq i_1 < i_2 < \cdots < i_r \leq n$.

Received February 1982; revised August 1982.

[1] Research supported by NSF Grant MCS 8201628.

[2] Research supported by Office of Naval Research under Contract N00014-77-C-0659.

AMS 1970 *subject classifications*. Primary 62E20; secondary 62G05, 62G10.

Key words and phrases. V-Statistics, U-statistics, Edgeworth expansion.

THEOREM 2.1. (a) *If for some integer* $s \geq 3$ *one has*

$$E_P|h(X_{j_1}, X_{j_2}, \cdots, X_{j_r})|^s < \infty \tag{2.3}$$

for all choices of $j_1, j_2, \cdots, j_r (1 \leq j_1, j_2, \cdots, j_r \leq r)$, *then the* pth *cumulant* $k_{p,n}(P)$ *of* V_n *under* P *is of the form*

$$k_{p,n}(P) = \sum_{m=p-1}^{s-1} n^{-m}\lambda_{m,p}(P) + o(n^{-s+1}), \qquad (2 \leq p \leq s). \tag{2.4}$$

The quantities $\lambda_{m,p}(P)$ *are independent of* n.

(b) *Suppose that, for some integer* $s \geq 3$ *one has*

$$E_P|h(X_1, X_2, \cdots, X_r)|^s < \infty. \tag{2.5}$$

Then the cumulants of the statistic U_n *also are of the form* (2.4).

PROOF. (a) Write

$$V_n = \int \cdots \int h(x_1, \cdots, x_r)F_n(dx_1) \cdots F_n(dx_r). \tag{2.6}$$

For $G = \sum_{i=1}^q \alpha_i\delta_{y_i}$ in $\mathscr{P}_f$, F_n may be expressed as $\sum_{i=1}^q \hat{\alpha}_i\delta_{y_i}$ (with $G^{\otimes n}$-probability one), where $\hat{\alpha}_i$ is the proportion of y_i's in the "sample" $\{X_1, \cdots, X_n\}$. Thus V_n becomes a polynomial in the q-variables $\hat{\alpha}_i$, $1 \leq i \leq q$. Hence by a result of James and Mayne (1962) (this may also be derived from the results of Leonov and Shiryaev, 1959), the pth cumulant of V_n under G is of the form

$$k_{p,n}(G) = \sum_{m=p-1}^{rp-1} n^{-m}\lambda_{m,p}(G), \qquad (p \geq 2). \tag{2.7}$$

On the other hand, for all G such that $E_G|V_n|^p < \infty$, one has (for all $n > rp$)

$$\begin{aligned} E_G V_n^p &= n^{-rp}E_G \int \cdots \int (\textstyle\prod_{t=1}^p h(x_{r(t-1)+1}, \cdots, x_{rt})) \\ &\quad \cdot (\delta_{X_1} + \delta_{X_2} + \cdots + \delta_{X_n})(dx_1) \cdots (\delta_{X_1} + \delta_{X_2} + \cdots + \delta_{X_n})(dx_{rp}) \\ &= n^{-rp} \int \cdots \int (\textstyle\prod_{t=1}^p h(x_{r(t-1)+1}, \cdots, x_{rt})) \\ &\quad \Big[\textstyle\sum_{m=1}^{rp} \frac{n!}{(n-m)!} \sum_2 \sum_1 E_G\{\delta_{X_1}(dx_{j_{11}})\delta_{X_1}(dx_{j_{12}}) \cdots \delta_{X_1}(dx_{j_{1s_1}}) \\ &\quad \delta_{X_2}(dx_{j_{21}}) \cdots \delta_{X_2}(dx_{j_{2s_2}}) \cdots \delta_{X_m}(dx_{j_{m1}}) \cdots \delta_{X_m}(dx_{j_{ms_m}})\}\Big]. \end{aligned} \tag{2.8}$$

Here, for a given m, $\sum_2$ denotes summation over all collections of m positive integers $\{s_1, s_2, \cdots, s_m\}$ satisfying $\sum s_i = rp$; for a given collection $\{s_1, s_2, \cdots, s_m\}$, $\sum_1$ denotes summation over all partitions of $\{1, 2, \cdots, rp\}$ into m groups of $s_1, s_2, \cdots, s_m$ elements, a typical partition being $(\{j_{11}, j_{12}, \cdots, j_{1s_1}\}, \{j_{21}, j_{22}, \cdots, j_{2s_2}\}, \{j_{m1}, j_{m2}, \cdots, j_{ms_m}\})$. Denote by $H_s(G)$ the distribution of the s-dimensional random vector $(X_1, X_1, \cdots, X_1)$ under G, and let $H_{s_1,s_2,\cdots,s_m}(G)$ stand for the measure

$$\begin{aligned} H_{s_1,s_2,\cdots,s_m}(G)(dx_1dx_2 \cdots dx_{rp}) &= \textstyle\sum_1 H_{s_1}(G)(dx_{j_{11}}dx_{j_{12}} \cdots dx_{j_{1s_1}})H_{s_2}(G) \\ &\quad \cdot (dx_{j_{21}}dx_{j_{22}} \cdots dx_{j_{2s_2}}) \cdots H_{s_m}(dx_{j_{m1}}dx_{j_{m2}} \cdots dx_{j_{ms_m}}). \end{aligned} \tag{2.9}$$

Also note that

$$\begin{aligned} n^{-rp}\frac{n!}{(n-m)!} &= n^{-rp}n(n-1) \cdots (n-m+1) \\ &= \textstyle\sum_{m'=1}^m (-1)^{m-m'}n^{-rp+m'}\theta(m-m'; m-1), \end{aligned} \tag{2.10}$$

where $\theta(i; N)$ is the sum of all products of i distinct integers taken from $\{1, 2, \cdots, N\}$,

$\theta(0; N) = 1$. From (2.8)–(2.10) one obtains

$$E_G V_n^p = \sum_{m=1}^{rp} \sum_{m'=1}^{m} (-1)^{m-m'} n^{-rp+m'} \theta(m-m'; m-1) \cdot \int \cdots \int \left(\prod_{t=1}^{p} h(x_{r(t-1)+1}, \cdots, x_{rt})\right) \sum\nolimits_2 H_{s_1,s_2,\cdots,s_m}(G)(dx_1 \cdots dx_{rp})$$

$$(2.11) \qquad = \sum_{j=0}^{rp-1} n^{-j} \Big\{ \sum_{m=rp-j}^{rp} (-1)^{m-rp+j} \theta(m-rp+j; m-1) \cdot \int \cdots \int \left(\prod_{t=1}^{p} h(x_{r(t-1)+1}, \cdots, x_{rt})\right) \sum\nolimits_2 H_{s_1,s_2,\cdots,s_m}(G)(dx_1 \cdots dx_{rp}) \Big\}$$

$$= \sum_{j=0}^{rp-1} n^{-j} \mu_{j,p}(G), \qquad (1 \le p \le s),$$

say. Here $\mu_{j,p}(G)$ is a linear combination (with coefficients not depending on n, G or h) of terms like

$$(2.12) \qquad \int \cdots \int \left(\prod_{t=1}^{p} h(x_{r(t-1)+1}, \cdots, x_{rt})\right) H_{s_1,s_2,\cdots,s_m}(G)(dx_1 \cdots dx_{rp}).$$

Using the familiar relations between moments and cumulants one has

$$(2.13) \qquad k_{p,n}(G) = \sum_{j=0}^{rp-1} n^{-j} \bar{\lambda}_{j,p}(G),$$

where $\bar{\lambda}_{j,p}(G)$ is a polynomial in $\mu_{j,p'} (1 \le p' \le p)$, whose coefficients are absolute constants. Since the map $G \to H_{s_1}(G)$ is continuous in the weak-star topology, so is the map $G \to H_{s_1,s_2,\cdots,s_m}(G)$. It follows that for a bounded continuous h the integral (2.12) is a weak-star continuous function of G; this implies that the maps $G \to \mu_{j,p}$ and, therefore, $G \to \bar{\lambda}_{m,p}(G)$ are continuous. If $p \ge 2$, then $\bar{\lambda}_{m,p}(G) = 0$ for $2 \le m < p-1$ and $G \in \mathscr{P}_f$. Also there exists $G_N \in \mathscr{P}_f (N = 1, 2, \cdots)$ such that G_N converges to P (This is where the separability of χ is made use of; see, e.g., Parthasarathy (1967), Theorem 6.3). Therefore, one must have $\bar{\lambda}_{m,p}(P) = 0$ for $1 \le m < p-1$. This completes the proof of (a) for bounded continuous h. Since functions of the form $\prod_{t=1}^{p} h(x_{r(t-1)+1}, \cdots, x_{rt})$ belonging to $L^1(\chi^{rp}, H_{s_1,s_2,\cdots,s_m}(P))$ may be approximated (in L^1) by continuous bounded functions of the same form, the proof is complete. Note that for this last argument (2.3) is needed.

(b) First assume (i) $h(x_1, x_2, \cdots, x_r) = 0$ *if* $x_i = x_j$ *for some* $i, j (i \ne j)$. Then the cumulants of U_n satisfy (2.4), since

$$U_n = \left(\frac{n!}{(n-r)!}\right)^{-1} n^r V_n = \left(\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right) \cdots \left(1-\frac{r-1}{n}\right)\right)^{-1} V_n = (1+o(1)) V_n.$$

Next, instead of (i) assume (ii) *P has no atoms*. Then modify h so as to satisfy (i); this does not change U_n, except on a set of probability zero. Finally, consider an arbitrary P. Let D be its set of atoms. Let D' be a subset of reals in one-one correspondence with D. Consider the space $\chi' = (\chi \backslash D) \cup R$, with $\chi \backslash D$ and R each carrying its own topology but their union is topologically disconnected. Then P lifted to this space χ' (by placing the discrete mass on D') is a weak-star limit of nonatomic probability measures. Extend h to $(\chi')^r$ by setting it zero if any coordinate is in $R \backslash D'$. Now apply an argument entirely analogous to that in the preceding paragraph.

REMARK 2.1.1. The U_n and V_n defined above are not centered around their expectations (under P). Centering has been avoided deliberately to ensure that h does not depend on P. For the general von Mises functional considered below centering seems unavoidable; this causes some technical problems.

REMARK 2.1.2. Under the hypotheses of Theorem 2.1 the pth cumulants of the normalized statistics $\sqrt{n}(V_n - EV_n)$, $\sqrt{n}(U_n - EU_n)$ are of the order $O(n^{-(p-2)/2})$, $2 \le p \le s$.

Let T be a *von Mises functional* defined on $\mathscr{P}_f \cup \{P\}$, and let the statistic $T(F_n)$ have the expansion

$$\begin{aligned}T(F_n) - T(G) &= \textstyle\sum_{i=1}^r \int \cdots \int T^{(i)}(G; x_1, x_2, \cdots, x_i) \prod_{j=1}^i (F_n - G)(dx_j) + R_n \\ &= \textstyle\sum_{i=1}^r V_{n,i}(G) + R_n \qquad (G \in \mathscr{P}_f \cup \{P\}),\end{aligned} \tag{2.14}$$

where $T^{(i)}$ is a real-valued, symmetric (in the arguments $x_1, \cdots, x_i$), Borel measurable function on $\mathscr{P}_f \cup \{P\} \times \chi^i$ satisfying

$$E_P|T^{(i)}(P; X_{j_1}, X_{j_2}, \cdots, X_{j_i})|^s < \infty, \qquad (1 \le i \le r), \tag{2.15}$$

for all $1 \le j_1, j_2, \cdots, j_i \le r$, and the "remainder term" R_n satisfies

$$E_P|R_n|^p = o(n^{-s+1}), \qquad (1 \le p \le s). \tag{2.16}$$

Write

$$V_n(G) = \textstyle\sum_{i=1}^r V_{n,i}(G). \tag{2.17}$$

Then

$$E_G V_n^p(G) = \textstyle\sum_3 V_{n,i_1}(G) V_{n,i_2}(G) \cdots V_{n,i_p}(G), \tag{2.18}$$

where $\sum_3$ denotes summation over all p-tuples $(i_1, i_2, \cdots, i_p)$ such that $1 \le i_1, \cdots, i_p \le r$. Now let $I_p = i_1 + \cdots i_p$ and write, as in (2.8),

$$\begin{aligned}E_G V_n^p(G) = \textstyle\sum_3 E_G \int \cdots \int &(\textstyle\prod_{t=1}^p T^{(i_t)}(G; x_{I_{t-1}+1}, \cdots, x_{I_t})) \\ &\cdot \Big[n^{-I_p} \textstyle\sum_{m=1}^{[I_p/2]} \dfrac{n!}{(n-m)!} \sum_2' \sum_1' (\delta_{X_1} - G)(dx_{j_{11}})(\delta_{X_1} - G)(dx_{j_{12}}) \\ &\cdots (\delta_{X_1} - G)(dx_{j_{1s_1}}) \cdots (\delta_{X_m} - G)(dx_{j_{m1}}) \cdots (\delta_{X_m} - G)(dx_{j_{ms_m}}) \Big].\end{aligned} \tag{2.19}$$

Here, for a given m, $\sum_2'$ denotes summation over all collections of m integers $\{s_1, s_2, \cdots, s_m\}$ satisfying $s_i \ge 2$ and $\sum s_i = I_p$; and $\sum_1'$ denotes, for each collection $\{s_1, s_2, \cdots, s_m\}$, summation over all partitions of $\{1, 2, \cdots, I_p\}$ into m subgroups of $s_1, s_2, \cdots, s_m$ elements such as $(\{j_{11}, j_{12}, \cdots, j_{1s_1}\}, \cdots, \{j_{m1}, j_{m2}, \cdots, j_{ms_m}\})$. Note that expectations of terms involving $s_i = 1$ for some i vanish. Next let $H_q(i_1, i_2, \cdots, i_\ell; G)$ denote the distribution of a q-dimensional random vector whose i_1th, $\cdots$, i_ℓth coordinates are X_1, while the remaining coordinates are i.i.d with distribution G and independent of X_1. Write $\tilde{H}_q(G)$ for the *signed* measure

$$\tilde{H}_q(G) = \textstyle\sum_{\ell=0}^q (-1)^{q-\ell} \sum_4 H_q(i_1, \cdots, i_\ell, G), \tag{2.20}$$

where $\sum_4$ denotes summation over all choices $\{i_1, i_2, \cdots, i_\ell\}$ of ℓ distinct integers from $\{1, 2, \cdots, q\}$. Now define

$$\begin{aligned}&\tilde{H}_{s_1, s_2, \cdots, s_m}(G; dx_1 dx_2 \cdots dx_{I_p}) \\ &\quad = \textstyle\sum_1' \tilde{H}_{s_1}(G)(dx_{j_{11}} dx_{j_{12}} \cdots dx_{j_{1s_1}}) \cdots \tilde{H}_{s_m}(G)(dx_{j_{m1}} dx_{j_{m2}} \cdots dx_{j_{ms_m}}).\end{aligned} \tag{2.21}$$

Then, as in (2.10),

$$\begin{aligned}E_G V_n^p(G) = \textstyle\sum_3 \Bigg[\sum_{j=\left[\frac{I_p+1}{2}\right]}^{I_p-1} n^{-j} \Big\{ &\textstyle\sum_{m=I_p-j}^{[I_p/2]} (-1)^{m-I_p+j} \theta(m - I_p + j; m-1) \\ &\cdot \int \cdots \int (\textstyle\prod_{t=1}^p T^{(i_t)}(G; x_{I_{t-1}+1}, \cdots, x_{I_t})) \sum_2' \tilde{H}_{s_1, \cdots, s_m}(G; dx_1 \cdots dx_{I_p}) \Big\} \Bigg].\end{aligned} \tag{2.22}$$

For $G = \sum_{i=1}^q \alpha_i \delta_{y_i}$, $V_n(G)$ is a polynomial in $\hat{\alpha}_i - \alpha_i$, so that the pth cumulant of $V_n(G)$ is of the order $O(n^{-p+1})$ under $G (2 \le p \le s)$. In view of (2.16) and (2.22), the proof of the following theorem is now complete.

THEOREM 2.2 *Suppose that* (2.14)–(2.16) *hold. Assume, in addition, that there exists a sequence* $\{G_N : N \geq 1\}$ *having finite support such that*

$$\lim_{N\to\infty} E_{G_N}(\textstyle\prod_{t=1}^{p} T^{(i_t)}(G_N; X_{t_1}, \cdots, X_{t_{i_t}})) = E_P(\prod_{t=1}^{p} T^{(i_t)}(P; X_{t_1}, \cdots, X_{t_{i_t}})) \tag{2.23}$$

for all $1 \leq i_1, i_2, \cdots, i_p \leq r$, *and all* $1 \leq t_1, t_2, \cdots, t_{i_t} \leq rp$ $(1 \leq t \leq p)$. *Then the* p*th cumulant of* $T(F_n)$ *under* P *is of the order* $O(n^{-p+1})$ *for* $2 \leq p \leq s$.

REMARK 2.2.1. Notice that the statement "condition (2.23) holds for *some* $\{G_N;\ N \geq 1\} \subset \mathscr{P}_f$" is much weaker than the statement "condition (2.23) holds for all sequences $\{G_N: N \geq 1\}$ converging to P (weak-star)", the latter being equivalent to saying that the integral is weak-star continuous at P (on $\mathscr{P}_f \cup \{P\}$). To illustrate this point, note that even such functionals as $T(G) = \int x^k G(dx)$, $k \geq 1$, are *not* weak-star continuous on $\mathscr{P}_f \cup \{P\}$, where P is a probability measure on the line having a finite kth moment. The difficulty is that one may place a mass $O(N^{-k/2})$ at $x = N$ which goes to zero to ensure weak-star convergence, but is large enough to blow up the integral as $N \to \infty$. On the other hand, one may integrate (with respect to P) a step-function approximation, $f_N(x)$ to x^k, which amounts to integrating x^k with respect to an appropriate $G_N \in \mathscr{P}_f$; and the latter integral $\int x^k G_N(dx)$ will converge to $\int x^k P(dx)$, as the intervals of constancy decrease to zero in width. These considerations apply to more general functions (see, Serfling (1980), pages 214–216, for examples).

REMARK 2.2.2. The fact that the sth cumulant of V_n (or T_n) is $O(n^{-s+1})$ when G has finite support means the vanishing of a number of polynomials in the variables $\mu_p(G)$. One should be able to prove that these polynomials are identically zero by showing that the $\mu_p(G)$'s assume a broad enough spectrum of values as G ranges over the set of all probability measures having finite support. This would enable one to dispense with the condition (2.23) in Theorem 2.2. However, we are unable to make this algebraic argument firm.

Finally, the method used here should be more widely applicable in deriving orders of magnitudes of cumulants.

3. A method of derivation of Edgeworth expansions of characteristic functions, and an unsolved problem.

In the present section we provide a method (which appears to be new) for the derivation of Cramér-Edgeworth expansions of characteristic functions of a class of statistics T_n having zero means, finite moment generating functions (m.g.f.'s), and cumulants $\chi_{p,n}$ satisfying

$$\chi_{p,n} = n^{-(p-2)/2}\lambda_p + o(n^{-(p-2)/2}), \qquad (p \geq 2), \qquad \lambda_2 > 0. \tag{3.1}$$

Let

$$f_n(\xi) = E \exp\{i\xi T_n\} \tag{3.2}$$

denote the characteristic function of T_n. One may write

$$f_n(\xi) = f(i\xi, \varepsilon), \tag{3.3}$$

with $\varepsilon = n^{-1/2}$. Under the additional assumption that $f(i\xi, \varepsilon)$ has an absolutely convergent power series expansion in ξ and ε in a neighborhood of the origin (0, 0), it is shown in Theorem 3.1 that $f_n(\xi)$ and its derivatives have a proper asymptotic expansion of the Cramér-Edgeworth type. The *unsolved problem* is to identify a large enough class of von Mises functionals for which this analyticity holds. In particular, we do not know if the analyticity property holds for U-statistics (see (2.2)) with kernels h satisfying:

$$E \exp\{th(X_1, X_2, \cdots, X_r)\} < \infty, \qquad (-\infty < t < \infty). \tag{3.4}$$

In remarks following the corollaries to Theorem 3.1 it is shown that the assumption of analyticity does hold for some special classes. We expect the moment computations of Section 2 to be crucial in resolving the problem of analyticity in the general case.

THEOREM 3.1. *Let $T_n (n = 1, 2, \cdots)$ be a sequence of random variables having zero means. Assume that* (i) *$E \exp\{tT_n\} < \infty$ for all $t(-\infty < t < \infty)$ and n,* (ii) *$f(i\xi, \varepsilon)$ can be extended as an analytic function $f(z, \eta)$ of the complex variables z and η in a neighborhood of the origin* (0, 0) *in $\mathbb{C}^2$, and* (iii) *the cumulants $\chi_{p,n}$ of T_n satisfy* (3.1). *Then the following results hold:*

(a) *There exist a positive constant δ_0 and polynomials P_j, whose coefficients do not depend on n, such that for all ξ, $-\delta_0\sqrt{n} < \xi < \delta_0\sqrt{n}$, one has*

$$f_n(\xi) = \exp\left\{-\frac{\lambda_2}{2}\xi^2\right\}\left(1 + \textstyle\sum_{j=1}^{\infty} n^{-j/2}P_j(i\xi)\right).$$

(b) *For every pair of integers m and p satisfying $p \geq 2$, $0 \leq m \leq p$, there exist positive constants δ_0, c_1, c_2 such that*

$$\left|\frac{d^m}{d\xi^m}\left[f_n(\xi) - \exp\left\{-\frac{\lambda_2}{2}\xi^2\right\}\left(1 + \textstyle\sum_{j=1}^{p-2} n^{-j/2}P_j(i\xi)\right)\right]\right|$$
$$\leq \frac{c_1}{n^{(p-1)/2}}\left[|\xi|^{p+1-m} + |\xi|^{3(p-1)+m}\right]\exp\{-c_2\xi^2\}, \qquad (|\xi| < \delta_0\sqrt{n}).$$

PROOF. Since $f(z, \eta)$ is analytic in a neighborhood of (0, 0), and $f(0, 0) = 1$, $\phi(z, \eta) = \log f(z, \eta)$ (we take the principal branch of the logarithm) is defined and analytic in a neighborhood of (0, 0). In view of (3.1) and the fact that $ET_n = 0$, one may express $\phi(z, \eta)$ as

$$\text{(3.5)} \qquad \begin{aligned}\phi(z, \eta) &= \frac{z^2}{2!}\left(\lambda_2 + \textstyle\sum_{j=1}^{\infty}\lambda_{2,j}\eta^j\right) + \cdots + \frac{z^k}{k!}\eta^{k-2}\left(\textstyle\sum_{j=0}^{\infty}\lambda_{k,j}\eta^j\right) + \cdots \\ &= z^2\left[\textstyle\sum_{k=2}^{\infty}\frac{(\eta z)^{k-2}}{k!}\left(\textstyle\sum_{j=0}^{\infty}\lambda_{k,j}\eta^j\right)\right], \qquad (\lambda_{k,0} = \lambda_k).\end{aligned}$$

Since this last series is absolutely convergent in a neighborhood of (0, 0), so is the series within square brackets. Let δ_1, δ_2 be two positive numbers such that this last series is absolutely convergent for $|z| = \delta_1$, $|\eta| = \delta_2$. Then

$$\text{(3.6)} \qquad \textstyle\sum_{k=2}^{\infty}\frac{(\delta_1\delta_2)^{k-2}}{k!}\sum_{j=0}^{\infty}|\lambda_{k,j}|\delta_2^j < \infty.$$

It follows that (3.5) is absolutely convergent for $|z\eta| \leq \delta_1\delta_2$ and $|\eta| \leq \delta_2$. Therefore, the last expression in (3.5) defines an analytic function in the region $D = \{(z, \eta) \in \mathbb{C}^2\colon |z| < \delta_1\delta_2/|\eta|, |\eta| < \delta_2\}$, and over this region $\exp\{\phi(z, \eta)\}$ defines an analytic continuation of $f(z, \eta)$. We shall refer to this extension also by $f(z, \eta)$. Since the characteristic function $\xi \to f_n(\xi)$ is entire (by assumption (i)) and since analytic continuations are unique, $f_n(\xi) = f(i\xi, n^{-1/2})$ for $-\infty < \xi < \infty$ (note that one could not assume a priori that this equality holds between f_n and the analytically extended f). In addition, on D one has

$$\text{(3.7)} \qquad |f(z, \eta) - 1| < c' < 1,$$

for some constant c', and $\phi(z, \eta)$ is the principal branch of the logarithm of $f(z, \eta)$ on D. The relations (3.5) now hold on D and one may rewrite the first relation in (3.5) as

$$\text{(3.8)} \qquad \log f(z, \eta) - \frac{\lambda_2}{2!}z^2 = \textstyle\sum_{j=1}^{\infty}\eta^j Q_j(z), \qquad (z, \eta) \in D,$$

where Q_j is a polynomial of degree $j + 2$. Thus

$$f(z, \eta) \exp\left\{-\frac{\lambda_2}{2} z^2\right\} = \exp\left\{\sum_{j=1}^{\infty} \eta^j Q_j(z)\right\} = 1 + \sum_{j=1}^{\infty} \eta^j P_j(z), \qquad (z, \eta) \in D, \tag{3.9}$$

where P_j's are appropriate polynomials. From (3.9) one gets

$$f(z, \eta) = \exp\left\{\frac{\lambda_2}{2} z^2\right\} (1 + \sum_{j=1}^{\infty} \eta^j P_j(z)), \qquad (z, \eta) \in D, \tag{3.10}$$

and, in particular (with $z = i\xi$, $\eta = n^{-1/2}$),

$$f_n(\xi) = \exp\left\{-\frac{\lambda_2}{2} \xi^2\right\} (1 + \sum_{j=1}^{\infty} n^{-j/2} P_j(i\xi)), \qquad (-\delta_1\delta_2 \sqrt{n} < \xi < \delta_1\delta_2 \sqrt{n}). \tag{3.11}$$

This proves part (a). To prove part (b) one may first approximate $\log f(z, \eta)$ by

$$\phi_p(z, \eta) = z^2 \sum_{k=2}^{p+2} \frac{(\eta z)^{k-2}}{k!} (\sum_{j=0}^{\infty} \lambda_{k,j} \eta^j). \tag{3.12}$$

Writing

$$\psi(z, \eta) = \phi(z, \eta) - \frac{\lambda_2}{2} z^2, \qquad \psi_p(z, \eta) = \phi_p(z, \eta) - \frac{\lambda_2}{2} z^2, \tag{3.13}$$

one has (using (3.6), or analyticity on D)

$$|\phi(z, \eta) - \phi_p(z, \eta)| = |\psi(z, \eta) - \psi_p(z, \eta)| \le c_3 |\eta|^{p-1} |z|^{p+1}, \qquad (z, \eta) \in D, \tag{3.14}$$

for an appropriate constant c_3. By (3.6) and (3.14), if δ_1 is small, then

$$|\exp\{\psi(z, \eta)\} - \exp\{\psi_p(z, \eta)\}| \le c_4 |\eta|^{p-1} |z|^{p+1} \exp\left\{\frac{\lambda_2 |z|^2}{4}\right\}, \tag{3.15}$$

for some constant c_4; this may be written as

$$\left|e^{-\frac{\lambda_2}{2} z^2} [f(z, \eta) - \exp\{\phi_p(z, \eta)\}]\right| \le c_4 |\eta|^{p-1} |z|^{p+1} \exp\left\{\frac{\lambda_2 |z|^2}{4}\right\}. \tag{3.16}$$

Letting $z = i\xi$, $\eta = n^{-1/2}$, (3.16) becomes

$$|f_n(\xi) - \exp\{\phi_p(i\xi, n^{-1/2})\}| \le c_4 n^{-(p-1)/2} |\xi|^{p+1} \exp\left\{-\frac{\lambda_2}{4} \xi^2\right\}, \qquad (|\xi| < \delta_1\delta_2 \sqrt{n}). \tag{3.17}$$

The comparison of $\exp\{\phi_p(i\xi, n^{-1/2})\}$ with $\exp\left\{-\frac{\lambda_2}{2} \xi^2\right\} (1 + \sum_{j=1}^{p-2} n^{-j/2} P_j(i\xi))$ is carried out exactly as in Lemmas 9.7, 9.8 in Bhattacharya and Ranga Rao (1976). □

COROLLARY 3.1.1. *Under the hypothesis of Theorem* 3.1 *one has the Berry-Esseen bound*

$$\sup_x |P(T_n \le x) - \Phi_{\lambda_2}(x)| \le c n^{-1/2}, \tag{3.18}$$

for some constant $c > 0$. *Here* Φ_{λ_2} *is the normal distribution function with mean zero and variance* λ_2.

PROOF. Use Theorem 3.1 (b) and Esseen's inequality (see Lemmas 12.1, 12.2 in Bhattacharya and Ranga Rao (1976)). □

COROLLARY 3.1.2. *Assume the hypothesis of Theorem* 3.1. *If, for some* $p \ge 2$, g *is a* p*-times continuously differentiable function on* $\mathbb{R}^1$ *such that* $\sup\{(1 + |x|^p)|g^{(m)}(x)|:$

$x \in I\!R^1\} < \infty$ *for* $0 \le m \le p$, *then*

$$(3.19) \qquad |Eg(T_n) - \int_{I\!R^1} g(x) \left[1 + \textstyle\sum_{j=1}^{p-2} n^{-j/2} P_j \left(-\frac{d}{dx} \right) \right] \phi_{\lambda_2}(x) dx | \le dn^{-(p-1)/2}$$

for some positive constant d.

PROOF. One may apply the method of Götze and Hipp (1978) to the estimate in Theorem 3.1 (b) to derive (3.19) directly. Alternatively, first establish (3.19) for the class of all Schwartz functions as in Bhattacharya and Ranga Rao (1976), Theorem 20.7, expressing the error estimate in terms of a Sobolev norm; then extend the result to a wider class by completion in the Sobolev norm. □

REMARK 3.1.3. Let $X_1, X_2, \cdots$ be an i.i.d. sequence having mean zero and a positive variance. The hypothesis of Theorem 3.1 is satisfied for the statistics $T_n = n^{-1/2}(X_1 + \cdots + X_n)$ if the m.g.f. of X_1 is finite everywhere. Of course, in this classical case Theorem 3.1 (b) holds under less stringent assumptions (see, e.g., Bhattacharya and Ranga Rao (1976), Chapter 2). Note, however, the conclusion of part (a) of Theorem 3.1 requires stronger assumptions than finiteness of moments.

REMARK 3.1.4 Let U_n be a U-statistic with kernel h (see (2.2)). Assume, without loss of generality, that $Eh(X_1, X_2, \cdots, X_r) = 0$. If $E \exp\{th(X_1, \cdots, X_r)\} < \infty$ for all t, $-\infty < t < \infty$, then hypothesis (i) of Theorem 3.1 is satisfied for the statistic $T_n = \sqrt{n}\ U_n$ (see Serfling (1980), Lemma C, page 200). In addition, assume $E\phi^2(X_1) = \lambda_2 > 0$, where $\phi(x) = Eh(x, X_2, \cdots, X_r)$. Then T_n is asymptotically normal (see Serfling (1980), Theorem A, page 192) and, by Theorem 2.1 (b), hypothesis (iii) of Theorem 3.1 also holds. It would be of great interest to see if hypothesis (ii) of Theorem 3.1 is a consequence of the above assumptions. We emphasize that this is the main unresolved problem in the context of the present article. For kernels h which are sums of products of functions of single variables, analyticity of $f(z, \eta)$ in a neighborhood of the origin in $\mathbb{C}^2$ has been proved by methods of statistical mechanics (see, e.g., Ruelle, 1969). However, for these special kernels an adequate theory of Edgeworth expansions has been derived in Bhattacharya and Ghosh (1978) under less stringent assumptions.

REMARK 3.1.5. Some partial expansions of characteristic functions of U-statistics have been obtained by Callaert, Janssen and Veraverbeke (1980).

Acknowledgement. The authors wish to thank the referee for suggesting that some terse comments in an earlier draft be expanded; the present Section 3 constitutes the resulting expansion. Thanks are also due to the Associate Editor for indicating that asymptotic expansions for von Mises functionals are of potential use in robust estimation. Finally, we are indebted to Charles M. Newman for pointing out that the analyticity hypothesis in Theorem 3.1 has been verified in the context of statistical mechanics for the special class of kernels mentioned at the end of Remark 3.1.4.

REFERENCES

BHATTACHARYA, R. N. (1977). Refinements of the multidimensional central limit theorem and applications. *Ann. Probab.* **5** 1–27.

BHATTACHARYA, R. N. and GHOSH, J. K. (1978). On the validity of the formal Edgeworth expansion. *Ann. Statist.* **6** 434–451.

BHATTACHARYA, R. N. and RANGA RAO, R. (1976). *Normal Approximation and Asymptotic Expansions.* Wiley, New York.

CALLAERT, H., JANSSEN, P. and VERAVERBEKE, N. (1980). An Edgeworth expansion for *U*-statistics. *Ann. Probab.* **8** 299–312.

GÖTZE, F. and HIPP, C. (1978). Asymptotic expansions in the central limit theorem under moment conditions. *Z. Wahrsch. verw. Gebiete* **42** 67–87.

JAMES, G. S. and MAYNE ALAN, J. (1962). Cumulants of functions of random variables. *Sankhya Ser. A* **24** 47–54.

LEONOV, V. P. and SHIRYAEV, A. N. (1959). On a method of calculation of semi-invariants. *Theor. Probab. Appl.*, **4** 319–329.

PARTHASARATHY, K. R. (1967). *Probability Measures on Metric Spaces.* Academic, New York.

RUELLE, D. (1969). *Statistical Mechanics: Rigorous Results.* Benjamin, Reading.

SERFLING, ROBERT J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

VON MISES, R. (1947). On the asymptotic distribution of differentiable statistical functions. *Ann. Math. Statist.* **18** 309–348.

WITHERS, C. S. (1980). The distribution and quantiles of a regular functional of the empirical distribution, Report No. 96, D.S.I.R., Wellington, New Zealand.

DEPARTMENT OF MATHEMATICS
SWAIN HALL EAST
INDIANA UNIVERSITY
BLOOMINGTON, INDIANA 47405