The Annals of Statistics
1973, Vol. 1, No. 6, 993-1026

ASYMPTOTIC EXPANSIONS RELATED TO MINIMUM CONTRAST ESTIMATORS

BY J. PFANZAGL

University of Cologne

This paper contains an Edgeworth-type expansion for the distribution of a minimum contrast estimator, and expansions suitable for the computation of critical regions of prescribed error (type one) as well as confidence intervals of prescribed confidence coefficient.

Furthermore, it is shown that, for one-sided alternatives, the test based on the maximum likelihood estimator as well as the test based on the derivative of the log-likelihood function is uniformly most powerful up to a term of order $O(n^{-1})$.

Finally, an estimator is proposed which is median unbiased up to an error of order $O(n^{-1})$ and which is—within the class of all estimators with this property—maximally concentrated about the true parameter up to a term of order $O(n^{-1})$.

The results of this paper refer to real parameters and to families of probability measures which are "continuous" in some appropriate sense (which excludes the common discrete distributions).

1. Introduction. It is well known that maximum likelihood estimators are asymptotically normally distributed. It is known as well that the accuracy of this normal approximation is insufficient for small sample sizes. A theorem like that in Pfanzagl (1971) stating that the error of the normal approximation is of order $O(n^{-\frac{1}{2}})$, is merely a more precise description of a desolate situation.

One possible expedient is the use of normalizing transformations. Approaches to a general theory of normalizing transformations can be found in the papers by Bol'shev (1959) and Borges (1971). In the case of hypotheses testing the use of normalizing transformations seems to be adequate, though it is not particularly flexible if an accuracy of higher order than $O(n^{-1})$ is wanted. In the theory of estimation it is of dubious value: Even if there is a normalizing transformation not depending on the unknown parameter, it leads to an estimator of some transform of the parameter, not of the parameter itself.

For these reasons we wish to explore another possibility, namely to give Edgeworth-type expansions for the distributions related to tests and estimators. This approach leads—under suitable regularity conditions—to approximations which are sufficiently accurate for all practical purposes.

The general theorem according to which minimum contrast estimators are asymptotically normally distributed up to an error of order $O(n^{-\frac{1}{2}})$ requires, of course, a number of regularity conditions (such as twofold differentiability of

Received September 1971; revised December 1972.

AMS subject classifications. 62E20, 62F10.

Key words and phrases. Maximum likelihood estimators, minimum contrast estimators, asymptotic expansions, tests, confidence intervals.

the densities with respect to the parameters, integrability of powers of these derivatives etc.). It holds, however, for "discrete" as well as "continuous" distributions. The existence of an Edgeworth expansion is, roughly speaking, bound to "continuous" distributions. More precisely, the distribution of $f^{(1)}(x, \theta)$ is to fulfill a condition similar to Cramér's Condition C. (Here $f^{(1)}$ denotes the first derivative of the contrast function.)

After the notations have been introduced in Section 2, the basic theorem containing the Edgeworth expansion for the minimum contrast estimator is given in Section 3. In Section 4 this result is applied to obtain a critical region of prescribed error (type one), and it is shown that this critical region, if based on the maximum likelihood estimator, is uniformly most powerful against one-sided alternatives up to a term of order $O(n^{-1})$. Finally, another critical region based on the log-likelihood function with the same optimum property is discussed. Corresponding results on confidence intervals are obtained in Section 5. In Section 6 an estimator is introduced which is median unbiased up to an error of $O(n^{-1})$ and which is—within the class of all estimators with this property—maximally concentrated about the true parameter up to a term of order $O(n^{-1})$. (The fact that an estimator with such a strong optimum property exists will doubtless enforce a reconsideration of the literature on BAN-estimators which are "best" at most up to a term $O(n^{-\frac{1}{2}})$, but hardly $O(n^{-1})$.)

Section 8 contains the precise statement of the regularity conditions and of the particular form of Condition C which is needed here.

In Section 7 it is shown that these assumptions are fulfilled for the most common location and scale parameter families (such as exponential, Cauchy and logistic), and the accuracy of the results is illustrated by a numerical example referring to the exponential distribution. The final Sections 9 and 10 contain lemmas and proofs of the theorems.

The following remarks should be kept in mind throughout the paper.

REMARK 1. The existence of asymptotic expansions depends on regularity conditions R_s, $s \geqq 3$, (involving properties of the derivatives $f^{(\nu)}(\cdot, \theta)$ of order $\nu = 1, \cdots, s$). Basically, to each $s \geqq 3$ there corresponds a different expansion of length $s - 1$. It follows easily, however, from the properties of asymptotic expansions that any expansion of length s simply "continues" the expansion of length $(s - 1)$ (i.e. the first $(s - 1)$ terms of the expansion of length s coincide with the expansion of length $s - 1$).

REMARK 2. Whenever in the following a relation is proved to hold up to an error term of order $o(n^{-(s-2)/2})$ under regularity conditions R_s, then the same relation holds true up to an error term of order $O(n^{-(s-1)/2})$ if regularity conditions R_{s+1} are fulfilled. This can be seen as follows: Under regularity conditions R_{s+1} we can add one more term to the asymptotic expansion, thus obtaining a relation which holds true up to an error term of order $o(n^{-(s-1)/2})$. Since the additional term is of order $O(n^{-(s-1)/2})$, the assertion follows. Since the validity of this argument can be easily verified in each particular case, we shall confine ourselves to this general remark.

REMARK 3. The case $s = 3$ is of special character in so far as the pertaining Edgeworth expansion of length 2 for the sum of independent identically distributed random variables is valid for arbitrary non-lattice distributions, and not only for distributions fulfilling Cramér's Condition C. This could be used to obtain asymptotic expansions like (3.1) and optimum assertions like (4.7), (5.4), and (6.5) for arbitrary non-lattice distributions. (Since we need uniformity with respect to θ and τ something slightly stronger than "non-lattice" has to be required. Probably, "non-lattice and $\sup_{\theta \in U_\theta} \sup_{|\tau-\theta|<a_\theta} E_\theta(1^{(1)}(\cdot, \theta, \tau)^4) < \infty$ for appropriate U_θ, a_θ" will do.) Since non-lattice distributions other than "continuous" ones are of limited practical interest, we shall confine ourselves to the case of "continuous" distributions in order not to overload the paper.

REMARK 4. The reader who wishes to apply the asymptotic expansions obtained in this paper (such as (3.1) or (4.2)) will be disappointed by the fact that the error term is only stated as $o(n^{-(s-2)/2})$ without precise bound. The technical reason for this deficiency is that our results are based on the Edgeworth expansion for the sum of independent identically distributed random variables, and that a precise error bound is not available even in this most simple case.

For the practician this is not a serious drawback. As common with asymptotic expansion we have to expect that any general error bound which might become available sooner or later will grossly overestimate the actual error.

Of greater practical relevance is the following problem: If the regularity conditions R_s are fulfilled, we have the choice between the expansions of length $1, 2, \cdots, s-1$. In the case of (3.1) for instance, to approximate $P^N\{\mathbf{x} \in X^N: n^{\frac{1}{2}}(\theta_n(\mathbf{x}) - \theta)/\beta(\theta) < t\}$ by $\Phi(t) + \varphi(t) \sum_{m=1}^{r} n^{-m/2} A_m(t, \theta)$ for $r = 0, 1, \cdots, s-2$.

If n is sufficiently large, the approximation for $r = s-2$ will be the most accurate one. What we would like to know, however, is something different, namely: Given t and the sample size n, which $r \in \{0, 1, \cdots, s-2\}$ renders the best approximation? This is not necessarily $r = s-2$.

REMARK 5. The generalization of our results to vector parameters is not at all straightforward. Recently, it was shown (see Pfanzagl (1973)) that for m.c. estimators of vector parameters the approximation by the normal distribution holds with an error of order $O(n^{-\frac{1}{2}})$ uniformly over the class of all convex sets. The practically relevant problem, however, is to give an Edgeworth expansion for the marginal distribution of the estimator for one component of the vector parameter, thus meeting the need of the most common situation, namely to make assertions about one (structural) parameter in the presence of nuisance parameters. In a recent paper, Chibisov (1972a) investigates the accuracy of the normal approximation for Neyman's optimal test and gives the first term of an asymptotic expansion (page 156, Theorem 2.1). His results are, however, not a vector-parameter analogue of our Proposition 1, since he does not investigate whether the optimality of Neyman's test against one-sided alternatives is of an order higher than $o(1)$.

2. Notations. Let $(X, \mathscr{A})$ be a measurable space and $P_\theta | \mathscr{A}$, $\theta \in \Theta$, a family of p-measures (probability measures), where $\Theta \subset \mathbb{R}$ is an open interval (possibly $\Theta = \mathbb{R}$). Let $\bar{\Theta}(\Theta^\circ)$ denote the closure (interior) of Θ in $\bar{\mathbb{R}} := [-\infty, +\infty]$ and $\mathscr{B}$ the Borel field over $\bar{\Theta}$.

For notational convenience we shall write $E_\theta(g)$ instead of $\int g(x)P_\theta(dx)$. For any measure $P|\mathscr{A}$ and any measurable function $g: X \to \mathbb{R}$ let $P * g | \mathscr{B}$ denote the *induced measure*, defined by $P * g(B) \equiv P(g^{-1}B)$, $B \in \mathscr{B}$.

A family of $\mathscr{A}$-measurable functions $f(\cdot, \theta): X \to \bar{\mathbb{R}}$, $\theta \in \bar{\Theta}$, is a family of *contrast functions* for $\{P_\theta : \theta \in \Theta\}$ if $E_\theta(f(\cdot, \tau))$ exists for all $\theta \in \Theta$, $\tau \in \bar{\Theta}$ and if

$$(2.1) \qquad E_\theta(f(\cdot, \theta)) < E_\theta(f(\cdot, \tau)) \qquad \text{for all} \quad \theta \in \Theta, \tau \in \bar{\Theta}, \theta \neq \tau .$$

Let $(X^{\mathrm{N}}, \mathscr{A}^{\mathrm{N}})$ be the countable Cartesian product of identical components $(X, \mathscr{A})$ and $P^{\mathrm{N}} | \mathscr{A}^{\mathrm{N}}$ the independent product of a countable number of identical components $P | \mathscr{A}$.

An *estimator* for the sample size n is an $\mathscr{A}^{\mathrm{N}}$-measurable map $\theta_n : X^{\mathrm{N}} \to \mathbb{R}$ which depends on $x_1, \cdots, x_n$ only.

A *minimum contrast* (m.c.) *estimator* for the sample size n is an estimator θ_n for which $\theta_n(X^{\mathrm{N}}) \subset \bar{\Theta}$ and

$$(2.2) \qquad \sum_{i=1}^n f(x_i, \theta_n(\mathbf{x})) = \inf_{\theta \in \bar{\Theta}} \sum_{i=1}^n f(x_i, \theta) .$$

We remark that the concept of m.c. estimators has been introduced by Huber (1967).

For any function $f(\cdot, \theta): X \to \bar{\mathbb{R}}$ let $f^{(i)}(\cdot, \theta)$ denote the ith derivative with respect to θ, i.e.,

$$f^{(i)}(x, \theta) \equiv \frac{\partial^i}{\partial \theta^i} f(x, \theta) , \qquad x \in X, i = 1, 2, \cdots .$$

Furthermore,

$$(2.3) \qquad \rho_{k,m}(\theta) \equiv E_\theta\left(\frac{\partial^m}{\partial \theta^m} (f^{(1)}(\cdot, \theta)^k)\right) .$$

Part of the results refers to the particular case that $P_\theta | \mathscr{A}$, $\theta \in \Theta$, is dominated by some σ-finite measure, say $\mu | \mathscr{A}$. In this case $p(\cdot, \theta)$ will be used to denote the density of $P_\theta | \mathscr{A}$ relative to $\mu | \mathscr{A}$, and $1(\cdot, \theta) \equiv \log p(\cdot, \theta)$. Any contrast function fulfilling

$$(2.4) \qquad f(x, \theta) = -1(x, \theta) \qquad \text{for all} \quad x \in X, \theta \in \Theta ,$$

will be called *likelihood contrast function*, and the pertaining m.c. estimator will be called *maximum likelihood* (m.l.) *estimator*. For maximum likelihood estimators it is useful to express our results by the functions

$$L_{ijk}(\theta) \equiv E_\theta((1^{(1)}(\cdot, \theta))^i (1^{(2)}(\cdot, \theta))^j (1^{(3)}(\cdot, \theta))^k)$$

rather than by $\rho_{k,m}$. Instead of L_{ij0} or L_{i00} we shall also write L_{ij} or L_i, respectively.

Notice that there are many different ways to express our results by the functions L_{ijk}, because these functions are not independent. Under suitable regularity conditions we have in particular

$$\begin{aligned} & L_{01} + L_2 = 0 \,, \\ (2.5) \qquad & L_{001} + 3L_{11} + L_3 = 0 \,, \\ & L_{0001} + 4L_{101} + 3L_{02} + 6L_{21} + L_4 = 0 \,; \end{aligned}$$

furthermore,

$$L_2' = 2L_{11} + L_3 \,, \qquad L_{11}' = L_{21} + L_{101} + L_{02} \,, \quad \text{etc.}$$

Finally, we shall use

$$\varphi(t) = (2\pi)^{-\frac{1}{2}} \exp[-\tfrac{1}{2}t^2]$$
$$\Phi(t) = \textstyle\int_{-\infty}^{t} \varphi(r)\, dr$$
$$N_\alpha \equiv \Phi^{-1}(\alpha) \,.$$

3. The basic theorem.

THEOREM 1. *Let θ_n, $n \in \mathbb{N}$, be a sequence of* m.c. *estimators. Assume that for $s \geqq 3$ the regularity conditions R_s are fulfilled and that Condition C_* is fulfilled for $h(\cdot, \theta, \tau) = f^{(1)}(\cdot, \tau)$, $\tau \in \Theta$.*

Then uniformly on compact subsets of Θ and uniformly for $t \in \mathbb{R}$:

$$(3.1) \qquad P_\theta^{\mathbb{N}}\left\{\mathbf{x} \in X^{\mathbb{N}} : n^{\frac{1}{2}} \frac{\theta_n(\mathbf{x}) - \theta}{\beta(\theta)} < t\right\} = \Phi(t) + \varphi(t) \textstyle\sum_{m=1}^{s-2} n^{-m/2} A_m(t, \theta) + o(n^{-(s-2)/2}) \,, \qquad \textit{where}$$

$$\beta = \rho_{20}^{\frac{1}{2}} \rho_{11}^{-1}$$

and $A_m(t, \theta)$ are polynomials in t with coefficients depending on θ. The coefficients remain bounded on compact subsets of θ.

We have in particular:

$$A_1(t, \cdot) = a_{10} + a_{11} t^2 \,, \qquad A_2(t, \cdot) = a_{20} t + a_{21} t^3 + a_{22} t^5$$

with

$$\begin{aligned} a_{10} &= -\tfrac{1}{6}\rho_{30}\rho_{20}^{-\frac{3}{2}} \\ a_{11} &= \tfrac{1}{6}\rho_{30}\rho_{20}^{-\frac{3}{2}} + \tfrac{1}{2}\rho_{12}\rho_{20}^{\frac{1}{2}}\rho_{11}^{-2} - \tfrac{1}{2}\rho_{21}\rho_{20}^{-\frac{1}{2}}\rho_{11}^{-1} \\ a_{20} &= \tfrac{1}{8} + \tfrac{1}{4}\rho_{21}\rho_{30}\rho_{20}^{-2}\rho_{11}^{-1} - \tfrac{1}{6}\rho_{31}\rho_{20}^{-1}\rho_{11}^{-1} - \tfrac{5}{24}\rho_{30}^2\rho_{20}^{-3} + \tfrac{1}{8}\rho_{40}\rho_{20}^{-2} \\ a_{21} &= \tfrac{1}{8} + \tfrac{1}{6}\rho_{13}\rho_{20}\rho_{11}^{-3} - \tfrac{1}{4}\rho_{12}\rho_{21}\rho_{11}^{-3} + \tfrac{3}{8}\rho_{21}^2\rho_{20}^{-1}\rho_{11}^{-2} \\ &\quad + \tfrac{1}{4}\rho_{30}\rho_{12}\rho_{20}^{-1}\rho_{11}^{-2} - \tfrac{1}{4}\rho_{22}\rho_{11}^{-2} - \tfrac{1}{2}\rho_{30}\rho_{21}\rho_{20}^{-2}\rho_{11}^{-1} \\ &\quad + \tfrac{1}{6}\rho_{31}\rho_{20}^{-1}\rho_{11}^{-1} + \tfrac{5}{36}\rho_{30}^2\rho_{20}^{-3} - \tfrac{1}{24}\rho_{40}\rho_{20}^{-2} \\ a_{22} &= -\tfrac{1}{72}\rho_{30}^2\rho_{20}^{-3} + \tfrac{1}{12}\rho_{30}\rho_{21}\rho_{20}^{-2}\rho_{11}^{-1} - \tfrac{1}{12}\rho_{30}\rho_{12}\rho_{20}^{-1}\rho_{11}^{-2} \\ &\quad - \tfrac{1}{8}\rho_{12}^2\rho_{20}\rho_{11}^{-4} + \tfrac{1}{4}\rho_{12}\rho_{21}\rho_{11}^{-3} - \tfrac{1}{8}\rho_{21}^2\rho_{20}^{-1}\rho_{11}^{-2} \,. \end{aligned}$$

If in the particular case of m.l. *estimators the relations* (2.5) *hold true then these formulas reduce to*

$$\begin{aligned}
\beta &= L_2^{-\frac{1}{2}} \\
a_{10} &= \tfrac{1}{6} L_2^{-\frac{3}{2}} L_3 \\
a_{11} &= L_2^{-\frac{3}{2}} (\tfrac{1}{3} L_3 + \tfrac{1}{2} L_{11}) \\
a_{20} &= \tfrac{1}{8} - \tfrac{1}{2} L_2^{-3} L_3 (\tfrac{5}{12} L_3 + L_{11}) + \tfrac{1}{2} L_2^{-2} (L_{21} + \tfrac{1}{4} L_4) \\
a_{21} &= \tfrac{1}{8} - L_2^{-3} L_3 (\tfrac{1}{9} L_3 + \tfrac{1}{4} L_{11}) + \tfrac{1}{2} L_2^{-2} (\tfrac{1}{3} L_{101} + L_{21} + \tfrac{1}{4} L_4) \\
a_{22} &= -\tfrac{1}{2} L_2^{-3} (\tfrac{1}{9} L_3^2 + \tfrac{1}{3} L_3 L_{11} + \tfrac{1}{4} L_{11}^2) \,.
\end{aligned}$$

A result similar to Theorem 1 has earlier been obtained by Mitrofanova (1967) for maximum likelihood estimators of vector parameters. Her paper contains, however, no explicit formulas for $A_m(t, \theta)$, and the proofs are rather sketchy—though perhaps not irremediably so. It remains, for instance, unclear, why R. R. Rao's Theorem 2 is applicable since (i)[1] the set A, defined on page 370, is hardly the inverse image of a convex set under the random variable z as defined on page 369 (notice that the functions H_i entering the definition of A are not even linear in z), (ii) some "continuity condition" is needed to ascertain that z fulfills Cramér's Condition C. But even the regularity conditions stated explicitly are rather restrictive; Mitrofanova's Condition 5, for instance, excludes the exponential family. Finally, the expansion is valid only for t in some interval depending somehow on θ. Recently Chibisov (1972b) announced an improved version of Mitrofanova's result for real parameters. We remark that the continuity condition assumed in Chibisov's theorem includes the derivatives $1^{(i)}(\cdot, \theta)$ for $i = 1, \cdots, s-2$ in contrast to our Condition C_* which is required for $1^{(1)}(\cdot, \theta)$ only.

In connection with Theorem 1 we should also mention the papers by Haldane and Smith (1956) as well as Shenton and Bowman (1963), where several cumulants of the distribution of the m.l. estimator are computed for discrete distributions. It is suggested that the knowledge of these cumulants could be used to reduce the bias of the m.l. estimator (Haldane and Smith (1956), page 101) or to fit a Pearson distribution (Kendall and Stuart (1961), Sections 18.19 and 18.20). Conditions under which this is admissible are missing. Presupposing that corresponding formulas hold for the cumulants of "continuous" distributions and neglecting the question of convergence one can compute the pertaining Edgeworth expansion. This yields the same coefficients a_{ij} as stated above for the m.l. estimator.

Since the coefficients of $A_m(t, \theta)$ in (3.1) are functions of the unknown parameter θ, formula (3.1) cannot be applied at once to improve the approximation for the distribution of the m.c. estimator. The observation that $A_1(t, \theta)$ is even in t, however, leads immediately to the result that for intervals symmetric about the unknown parameter the normal approximation is particularly good.

[1] This was brought to my attention by Professor Bickel.

COROLLARY 1. *Let* θ_n, $n \in \mathbb{N}$, *be a sequence of* m.c. *estimators. Assume that regularity conditions* R_3 *are fulfilled and that Condition* C_* *is fulfilled for* $h(\cdot, \theta, \tau) = f^{(1)}(\cdot, \tau)$, $\tau \in \Theta$.

Then uniformly on compact subsets of Θ *and uniformly for* $t \in \mathbb{R}$

$$P_\theta^{\mathrm{N}}\left\{\mathbf{x} \in X^{\mathrm{N}} : n^{\frac{1}{2}} \frac{|\theta_n(\mathbf{x}) - \theta|}{\beta(\theta)} < t\right\} = \int_{-t}^{+t} (2\pi)^{-\frac{1}{2}} \exp[-r^2/2]\, dr + o(n^{-\frac{1}{2}}) . \tag{3.2}$$

4. Tests. Theorem 1 can be applied to obtain for the hypothesis θ_0 a critical region with error of type one equal to $\alpha + o(n^{-(s-2)/2})$. Since

$$t \to \Phi(t) + \varphi(t) \sum_{m=1}^{s-2} n^{-m/2} A_m(t, \theta_0)$$

is a continuous function which approaches 0 for $t \to -\infty$ and 1 for $t \to \infty$, for every $\alpha \in (0, 1)$ there exists $t_{n,\theta_0}(\alpha)$ such that

$$\Phi(t_{n,\theta_0}(\alpha)) + \varphi(t_{n,\theta_0}(\alpha)) \sum_{m=1}^{s-2} n^{-m/2} A_m(t_{n,\theta_0}(\alpha), \theta_0) = 1 - \alpha . \tag{4.1}$$

Together with (3.1) this implies

$$P_{\theta_0}^{\mathrm{N}}\{\mathbf{x} \in X^{\mathrm{N}} : \theta_n(\mathbf{x}) \geqq \theta_0 + n^{-\frac{1}{2}}\beta(\theta_0) t_{n,\theta_0}(\alpha)\} = \alpha + o(n^{-(s-2)/2}) .$$

For practical purposes it will be convenient to avoid the solution of equation (4.1) and to give the lower bound of the critical region explicitly.

THEOREM 2. *Let* θ_n, $n \in \mathbb{N}$, *be a sequence of* m.c. *estimators. Assume that for some* $s \geqq 3$ *the regularity conditions* R_s *are fulfilled, and that Condition* C_* *is fulfilled for* $h(\cdot, \theta, \tau) = f^{(1)}(\cdot, \tau)$, $\tau \in \Theta$.

Then there exists a function

$$G_{n,\alpha}^{(s)}(\theta) \equiv \theta - n^{-\frac{1}{2}} N_\alpha \beta(\theta) + \sum_{m=1}^{s-2} n^{-(m+1)/2} B_m(-N_\alpha, \theta) \tag{4.2}$$

(*where* $B_m(t, \theta)$ *is a polynomial in* t *with coefficients depending on* θ) *such that*

$$\{\mathbf{x} \in X^{\mathrm{N}} : \theta_n(\mathbf{x}) \geqq G_{n,\alpha}^{(s)}(\theta)\} \tag{4.3}$$

defines a critical region for the hypothesis θ *with error of type one equal to* $\alpha + o(n^{-(s-2)/2})$, *uniformly on compact subsets of* Θ.

We have in particular

$$B_1(t, \cdot) = b_{10} + b_{11} t^2 , \qquad B_2(t, \cdot) = b_{20} t + b_{21} t^3$$

with

$$b_{10} = \tfrac{1}{6}\rho_{30}\rho_{20}^{-1}\rho_{11}^{-1}$$

$$b_{11} = (-\tfrac{1}{6}\rho_{30}\rho_{20}^{-1} - \tfrac{1}{2}\rho_{12}\rho_{20}\rho_{11}^{-2} + \tfrac{1}{2}\rho_{21}\rho_{11}^{-1})\rho_{11}^{-1}$$

$$\begin{aligned} b_{20} = (&-\tfrac{1}{8} - \tfrac{1}{6}\rho_{12}\rho_{20}^{-1}\rho_{11}^{-2}\rho_{30} - \tfrac{1}{12}\rho_{21}\rho_{30}\rho_{20}^{-2}\rho_{11}^{-1} + \tfrac{1}{6}\rho_{31}\rho_{20}^{-1}\rho_{11}^{-1} \\ &+ \tfrac{5}{36}\rho_{30}^2\rho_{20}^{-3} - \tfrac{1}{8}\rho_{40}\rho_{20}^{-2})\rho_{20}^{\frac{1}{2}}\rho_{11}^{-1} \end{aligned}$$

$$\begin{aligned} b_{21} = (&-\tfrac{1}{8} - \tfrac{1}{18}\rho_{30}^2\rho_{20}^{-3} + \tfrac{1}{2}\rho_{12}^2\rho_{20}\rho_{11}^{-4} + \tfrac{1}{8}\rho_{21}^2\rho_{20}^{-1}\rho_{11}^{-2} + \tfrac{1}{6}\rho_{12}\rho_{20}^{-1}\rho_{11}^{-2}\rho_{30} \\ &+ \tfrac{1}{12}\rho_{21}\rho_{20}^{-2}\rho_{11}^{-1}\rho_{30} - \tfrac{3}{4}\rho_{12}\rho_{21}\rho_{11}^{-3} - \tfrac{1}{6}\rho_{13}\rho_{20}\rho_{11}^{-3} \\ &+ \tfrac{1}{4}\rho_{22}\rho_{11}^{-2} - \tfrac{1}{6}\rho_{31}\rho_{20}^{-1}\rho_{11}^{-1} + \tfrac{1}{24}\rho_{40}\rho_{20}^{-2})\rho_{20}^{\frac{1}{2}}\rho_{11}^{-1} . \end{aligned}$$

If in the particular case of m.l. *estimators the relations* (2.5) *hold true, then these formulas reduce to*

$$b_{10} = -\tfrac{1}{6}L_2^{-2}L_3$$
$$b_{11} = -L_2^{-2}(\tfrac{1}{3}L_3 + \tfrac{1}{2}L_{11})$$
$$b_{20} = -\tfrac{1}{8}L_2^{-\frac{1}{2}} + \tfrac{1}{3}L_2^{-\frac{7}{2}}L_3(\tfrac{1}{1}\tfrac{1}{2}L_3 + 2L_{11}) - \tfrac{1}{2}L_2^{-\frac{5}{2}}(L_{21} + \tfrac{1}{4}L_4)$$
$$b_{21} = -\tfrac{1}{8}L_2^{-\frac{1}{2}} + \tfrac{1}{2}L_2^{-\frac{7}{2}}(\tfrac{5}{9}L_3^2 + \tfrac{5}{3}L_3L_{11} + L_{11}^2) - \tfrac{1}{2}L_2^{-\frac{5}{2}}(\tfrac{1}{3}L_{101} + L_{21} + \tfrac{1}{4}L_4)\,.$$

In the following we shall state an optimum property for the critical region specified in (4.3).

THEOREM 3. (i) *Assume that the likelihood contrast function fulfills regularity conditions* (iv), (v a), *and* (2.5). *Assume that condition* (xi) *is fulfilled and that* $1^{(1)}$ *fulfills condition* L_4, $1^{(2)}$ *fulfills condition* L_3. *Assume, furthermore, that Condition* C_* *is fulfilled for*

$$\begin{aligned} h(\cdot, \theta, \tau) &= (\theta - \tau)^{-1}(1(\cdot, \theta) - 1(\cdot, \tau)) & \theta \neq \tau\,, \\ &= 1^{(1)}(\cdot, \tau) & \theta = \tau \quad \theta, \tau \in \Theta\,. \end{aligned}$$

Assume, finally, that for every $n \in \mathbb{N}$, $\theta \in \Theta$, *we are given an* $\mathscr{A}^n$*-measurable critical function* $\varphi_{n,\theta}: X^{\mathrm{N}} \to [0, 1]$ *such that for some* $\alpha \in (0, 1)$ *and some compact* $K \subset \Theta$ *we have uniformly for* $\theta \in K$:

$$E_\theta^{\mathrm{N}}(\varphi_{n,\theta}) = \alpha + o(n^{-\frac{1}{2}})\,. \tag{4.4}$$

Then uniformly for $\theta \in K$, $t \in \mathbb{R}$

$$E^{\mathrm{N}}_{\theta+n^{-1/2}t}(\varphi_{n,\theta}) \begin{Bmatrix} \leqq \\ \geqq \end{Bmatrix} H_{n,\alpha}(t, \theta) + o(n^{-\frac{1}{2}}) \qquad \textit{if} \quad t \begin{Bmatrix} > \\ < \end{Bmatrix} 0 \tag{4.5}$$

with

$$\begin{aligned} H_{n,\alpha}(t, \theta) = {} & \Phi(N_\alpha + tL_2(\theta)^{\frac{1}{2}}) \\ & + \varphi(N_\alpha + tL_2(\theta)^{\frac{1}{2}})n^{-\frac{1}{2}}\tfrac{1}{6}tL_2(\theta)^{-\frac{1}{2}} \\ & \times [t(3L_{11}(\theta) + L_3(\theta)) - N_\alpha L_2(\theta)^{-\frac{1}{2}}L_3(\theta)]\,. \end{aligned} \tag{4.6}$$

(ii) *Assume that the regularity conditions specified in Theorem 2 are fulfilled for the likelihood contrast function, together with condition* (xi), (2.5), *and condition* L_4 *for* $1^{(1)}$ *and condition* L_3 *for* $1^{(2)}$. *Assume, finally, that condition* C_* *is fulfilled for* $h(\cdot, \theta, \tau) = 1^{(1)}(\cdot, \tau)$, $\tau \in \bar{\Theta}$. *Then equality holds in* (4.5) *if* $\varphi_{n,\theta}$ *is the indicator function of the critical region defined by* (4.3) *for some* $s \geqq 3$, *provided it is based on a* m.l. *estimator. The equality holds uniformly for* $\theta \in K$, $|t| \leqq c_K n^{\frac{1}{2}}$.

(iii) *Assume that the regularity conditions specified in* (i) *and* (ii) *are fulfilled. Then the following relation holds uniformly for* $\theta \in K$, $|t| \leqq c_K n^{\frac{1}{2}}$, *provided* θ_n *is a* m.l. *estimator.*

$$E^{\mathrm{N}}_{\theta+n^{-1/2}t}(\varphi_{\theta,n}) \begin{Bmatrix} \leqq \\ \geqq \end{Bmatrix} P^{\mathrm{N}}_{\theta+n^{-1/2}t}\{\mathbf{x} \in X^{\mathrm{N}} : \theta_n(\mathbf{x}) \geqq G^{(s)}_{n,\alpha}(\theta)\} + o(n^{-\frac{1}{2}}) \qquad \textit{for} \quad t \begin{Bmatrix} > \\ < \end{Bmatrix} 0\,. \tag{4.7}$$

(4.5) provides an approximate upper bound for the power function of any critical function with error of type one equal to $\alpha + o(n^{-\frac{1}{2}})$. In other words, $t \to H_{n,\alpha}(t, \theta)$ is an approximation to the envelope power function. By the same arguments one can obtain in (4.5) an approximation up to $o(n^{-(s-2)2})$ if (4.4) holds with $o(n^{-\frac{1}{2}})$ replaced by $o(n^{-(s-2)/2})$, and if further terms are added to $H_{n,\alpha}$ in (4.6), provided that appropriate regularity conditions are fulfilled. We shall abstain from doing so, because we need (4.5) mainly for the proof of the optimum property (4.7).

Though it is not uncommon to base the comparison of tests on a sequence of alternatives of the type $\theta + n^{-\frac{1}{2}}t$, it seems not useless to give an operational justification for this approach. The use of alternatives $\theta + n^{-\frac{1}{2}}t$ is appropriate if the experimenter is able to specify which deviations from the hypothesis θ (say values $\tau > \theta + \Delta$) are practically relevant and should therefore be rejected with high probability (say .99). In this case he will adjust the sample size n to the particular problem in question by choosing n so that $E_\theta{}^{\mathrm{N}}(\varphi_{n,\theta}) = .01$ and $E^{\mathrm{N}}_{\theta+\Delta}(\varphi_{n,\theta}) = .99$. If $\varphi_{n,\theta}$ is derived from some estimator T_n which is normally distributed with asymptotic mean θ and asymptotic variance $\sigma(\theta)^2$, this will be the case if $n^{-\frac{1}{2}}$ is related to Δ by $\Delta = 4, 6\,\sigma(\theta)n^{-\frac{1}{2}}$. In other words: The sample sizes used by an experimenter will vary according to the problem. If they are adjusted to the corresponding problem, it will always be the behavior of the test for alternatives $\theta + tn^{-\frac{1}{2}}$ with t somewhere between $3\sigma(\theta)$ and $6\sigma(\theta)$ which is of interest.

Using regularity conditions R_2 only, the same pattern of proof shows that (4.7) holds with $o(n^{-\frac{1}{2}})$ replaced by $O(n^{-\frac{1}{2}})$ without any conditions like "non-lattice" or "C". This improves a result of Wald (1941 a) page 10, Theorem 1, where (4.7) with $o(1)$ was established. (See also C. R. Rao, (1962), page 54, Theorem 2.) Hence our Theorem 3 sharpens a well-known optimum property of the m.l. estimator for the particular case of "continuous" distributions.

According to Remark 2 in Section 1, (4.7) holds true even with $o(n^{-\frac{1}{2}})$ replaced by $O(n^{-1})$ under additional regularity conditions. A somewhat tedious computation shows that (4.7) holds not true any more with $o(n^{-\frac{1}{2}})$ replaced by $o(n^{-1})$, even if the regularity conditions are appropriately strengthened.

Using the concept of deficiency introduced by Hodges and Lehmann (1970) these results (as well as the results of Proposition 1, Theorem 5 and Theorem 6 below) can be given a more concise formulation: Let $\varphi_n{}^{(i)}$, $n \in \mathbb{N}$, $i = 1, 2$, be two sequences of critical functions with $E_\theta{}^{\mathrm{N}}(\varphi_n{}^{(i)}) = \alpha \in (0, 1)$ for $n \in \mathbb{N}$, $i = 1, 2$. If t is given, let $m(n)$ be the smallest integer m such that $E^{\mathrm{N}}_{\theta+n^{-\frac{1}{2}}t}(\varphi_m{}^{(2)}) \geqq E^{\mathrm{N}}_{\theta+n^{-\frac{1}{2}}t}(\varphi_n{}^{(1)})$. If $m(n)/n \to 1$ (i.e. if the asymptotic efficiency of $\varphi_n{}^{(2)}$ relative to $\varphi_n{}^{(1)}$ is 1), the deficiency of $\varphi_n{}^{(2)}$ relative to $\varphi_n{}^{(1)}$ is defined as $m(n) - n$. In typical cases we have

$$E^{\mathrm{N}}_{\theta+n^{-\frac{1}{2}}t}(\varphi_n{}^{(i)}) = a_i(t, \theta) + n^{-\frac{1}{2}}b_i(t, \theta) + n^{-1}c_i(t, \theta) + o(n^{-1}), \qquad i = 1, 2.$$

We have $m(n)/n \to 1$ iff $a_1(t, \theta) = a_2(t, \theta)$. Under this assumption we have

$m(n) - n = O(n^{\frac{1}{2}})$ in general and $m(n) - n = O(1)$ iff $b_1(t, \theta) = b_2(t, \theta)$. The asymptotic deficiency is zero iff $c_1(t, \theta) = c_2(t, \theta)$.

Hence we may summarize our results as follows: The asymptotic deficiency of the test based on the m.l. estimator as compared to the most powerful test is $O(1)$ (depending on t), but not zero, in general. The surprising fact is that it is not $O(n^{\frac{1}{2}})$. It would be interesting to know whether there are one-sided tests which are approximately most powerful in the sense that they have asymptotic deficiency zero (as compared to the most powerful test) for all $t > 0$. Regrettably, neither the test based on the m.l. estimator nor the test described in Proposition 1 has this property. This is the point to remember that there is one important type of families of p-measures for which, for any sample size n, $(x_1, \cdots, x_n) \rightarrow \sum_{i=1}^n (1(x_i, \theta + n^{-\frac{1}{2}}t) - 1(x_i, \theta))$ is a monotone function of the m.l. estimator, so that the test based on the m.l. estimator is uniformly most powerful for families of this type: These are the exponential families (see Huzurbazar (1947)). Perhaps the exponential families are the only ones admitting tests with asymptotic deficiency equal to zero for all $t > 0$.

Since the m.l. estimator is difficult to compute in some cases, it seems worthwhile to mention that the critical region based on $\sum_{i=1}^n 1^{(1)}(x_i, \theta)$ has the same optimum property as the critical region based on the m.l. estimator.

PROPOSITION 1. *Assume that the likelihood contrast function fulfills regularity conditions* (iv), (v a). *Assume, furthermore, that conditions* (xi), *and* (2.5) *are fulfilled and that* $1^{(1)}$ *fulfills condition* L_4, $1^{(2)}$ *fulfills condition* L_3. *Assume, finally, that Condition* C_* *is fulfilled for* $h(\cdot, \theta, \tau) \equiv 1^{(1)}(\cdot, \theta)$, $\theta \in \bar{\Theta}$. *Then the critical region*

$$\{\mathbf{x} \in X^{\mathrm{N}}: n^{-\frac{1}{2}} \sum_{i=1}^n 1^{(1)}(x_i, \theta) \geqq -N_\alpha L_2(\theta)^{\frac{1}{2}} + n^{-\frac{1}{2}} L_2(\theta)^{-1} L_3(\theta) \tfrac{1}{6}(N_\alpha{}^2 - 1)\}$$

has the power function $H_{n,\alpha}(t, \theta)$ *given by* (4.6) (*and hence in particular an error of type one equal to* $\alpha + o(n^{-\frac{1}{2}})$), *uniformly on compact subsets of* θ *and* $|t| \leqq c_K n^{\frac{1}{2}}$.

That a critical region based on $\sum_{i=1}^n 1^{(1)}(x_i, \theta)$ is optimal up to $o(1)$ is to be expected on account of Theorem 3 (ii) and the the fact that $(n^{\frac{1}{2}}[(\theta_n(\mathbf{x}) - \theta)/\beta(\theta)] - n^{-\frac{1}{2}}\beta(\theta) \sum_1^n 1^{(1)}(x_i, \theta)_{n \in \mathrm{N}}$ converges to zero in probability. It was proved by Wald (1941 b), page 403, Theorem 1. A more informal statement of Wald's result occurs in C. R. Rao and Poti (1964, page 439). Proposition 1 asserts optimality up to $o(n^{-\frac{1}{2}})$ (and even $O(n^{-1})$ under appropriate regularity conditions).

5. Confidence intervals. Theorem 2 leads in the usual way to a confidence set. From the critical region $\{\mathbf{x} \in X^{\mathrm{N}}: \theta_n(\mathbf{x}) \geqq G_{n,\alpha}^{(s)}(\theta)\}$ with error of type one equal to $\alpha + o(n^{-(s-2)/2})$ we obtain the confidence set $C_n(\mathbf{x}) \equiv \{\theta \in \Theta: \theta_n(\mathbf{x}) < G_{n,\alpha}^{(s)}(\theta)\}$ with the confidence coefficient $P_\theta{}^{\mathrm{N}}\{\mathbf{x} \in X^{\mathrm{N}}: \theta \in C_n(\mathbf{x})\} = 1 - \alpha + o(n^{-(s-2)/2})$. Since the set $C_n(\mathbf{x})$ is not necessarily an interval, it seems preferable to apply the confidence interval specified by (5.1) which shares the optimum property of $C_n(\mathbf{x})$ resulting from Theorem 3.

THEOREM 4. *Let θ_n be a sequence of* m.c. *estimators. Assume that the conditions specified in Theorem* 2 *are fulfilled, that b_{ij} (as defined in Theorem* 2) *is* $(s-2-i)$-*times differentiable and that the derivative of order $s-2-i$ fulfills a Lipschitz-condition. Assume, furthermore, that β is* $(s-2)$-*times differentiable and that the derivative of order $s-2$ fulfills a Lipschitz-condition. Then for every $n \in \mathbb{N}$, $\alpha \in (0, 1)$ there exists a function*

$$F_{n,\alpha}^{(s)}(\theta) \equiv \theta + n^{-\frac{1}{2}} N_\alpha \beta(\theta) + \sum_{m=1}^{s-2} n^{-(m+1)/2} B_m^*(-N_\alpha, \theta)$$

(where $B_m^*(t, \theta)$ is a polynomial in t with coefficients depending on θ) such that

$$(F_{n,\alpha}^{(s)}(\theta_n(\mathbf{x})), \infty) \tag{5.1}$$

is a confidence interval with confidence coefficient equal to $1 - \alpha + o(n^{-(s-2)/2})$, *uniformly on compact subsets of* Θ.

We have in particular

$$B_1^*(t, \cdot) = b_{10}^* + b_{11}^* t^2, \qquad B_2^*(t, \cdot) = b_{20}^* t + b_{22}^* t^3$$

with

$$\begin{aligned} b_{10}^* &= -b_{10} \\ b_{11}^* &= -b_{11} + \tfrac{1}{2}(\beta^2)' \\ b_{20}^* &= -b_{20} + (\beta b_{10})' \\ b_{21}^* &= -b_{21} + (\beta b_{11} - \tfrac{1}{2}\beta^2\beta')' \end{aligned}$$

(where $'$ denotes the derivative with respect to θ).

If in the particular case of m.l. *estimators the relations* (2.5) *hold true, then these formulas reduce to*

$$\begin{aligned} b_{10}^* &= \tfrac{1}{6}L_2^{-2}L_3 \\ b_{11}^* &= -\tfrac{1}{6}L_2^{-2}(L_3 + 3L_{11}) \\ b_{20}^* &= \tfrac{1}{18}L_2^{-\frac{7}{2}}L_3(2L_3 + 3L_{11}) - \tfrac{1}{24}L_2^{-\frac{5}{2}}L_4 + \tfrac{1}{8}L_2^{-\frac{1}{2}} \\ b_{21}^* &= -\tfrac{1}{72}L_2^{-\frac{7}{2}}(5L_3^2 + 30L_{11}L_3 + 36L_{11}^2) + \tfrac{1}{24}L_2^{-\frac{5}{2}}(L_4 + 6L_{21} + 4L_{101}) + \tfrac{1}{8}L_2^{-\frac{1}{2}}. \end{aligned}$$

THEOREM 5. (i) *Assume that the regularity conditions specified in Theorem* 3 (i) *are fulfilled. Assume that for every $n \in \mathbb{N}$, $\mathbf{x} \in X^{\mathbb{N}}$, we are given a confidence set $C_n(\mathbf{x})$ such that $\{\mathbf{x} \in X^{\mathbb{N}} : \theta \in C_n(\mathbf{x})\} \in \mathscr{A}^{\mathbb{N}}$ for every $\theta \in \Theta$ and for some $\alpha \in (0, 1)$ and some compact $K \subset \Theta$ we have uniformly for $\theta \in K$:*

$$P_\theta^{\mathbb{N}}\{\mathbf{x} \in X^{\mathbb{N}} : \theta \in C_n(\mathbf{x})\} = 1 - \alpha + o(n^{-\frac{1}{2}}). \tag{5.2}$$

Then uniformly for $\theta \in K$, $t \in \mathbb{R}$

$$P_\theta^{\mathbb{N}}\{\mathbf{x} \in X^{\mathbb{N}} : \theta - n^{-\frac{1}{2}}t \in C_n(\mathbf{x})\} \tag{5.3}$$
$$\begin{Bmatrix} \geqq \\ \leqq \end{Bmatrix} 1 - H_{n,\alpha}(t, \theta - n^{-\frac{1}{2}}t) + o(n^{-\frac{1}{2}}) \qquad \text{if} \quad t \begin{Bmatrix} > \\ < \end{Bmatrix} 0$$

(where $H_{n,\alpha}$ is defined by (4.6)).

(ii) *If the regularity conditions specified in Theorem* 3 (ii) *and Theorem* 4 *are fulfilled for the likelihood contrast function, then equality holds in* (5.3) *if C_n is the confidence interval defined by* (5.1) *for some $s \geqq 3$, provided this confidence interval is based on a* m.l. *estimator.*

(iii) *If the regularity conditions specified in* (i) *and* (ii) *are fulfilled, then the following relation holds uniformly for* $\theta \in K$, $|t| \leqq c_K n^{\frac{1}{2}}$, *provided* θ_n *is a* m.l. *estimator*:

$$(5.4)\qquad P_\theta^{\mathrm{N}}\{\mathbf{x} \in X^{\mathrm{N}} : \theta - n^{-\frac{1}{2}}t \in C_n(\mathbf{x})\}$$
$$\begin{Bmatrix}\geqq\\ \leqq\end{Bmatrix} P_\theta^{\mathrm{N}}\{\mathbf{x} \in X^{\mathrm{N}} : F_{n,\alpha}^{(s)}(\theta_n(\mathbf{x})) \leqq \theta - n^{-\frac{1}{2}}t\} + o(n^{-\frac{1}{2}}) \qquad \text{if} \quad t \begin{Bmatrix}>\\ <\end{Bmatrix} 0\,.$$

We remark that an optimum property corresponding to (5.4) with $o(n^{-\frac{1}{2}})$ replaced by $o(n^{-1})$ does not hold true any more.

Since the accuracy of the normal approximation has always been felt insufficient, several proposals have been made to obtain confidence sets, the confidence coefficient of which is in better agreement with the prescribed value. For m.l. estimators, Welch (1965, page 3, formula (22)) gives the expression for $F_{n,\alpha}^{(3)}$. He arrives at this result by determining the cumulants of $n^{\frac{1}{2}}(\theta_n - \theta)/\beta(\theta)$ from a formal expansion and applying then the Cornish–Fisher formula for percentage points. Working with formal expansions only, Welch overlooks that the use of $F_{n,\alpha}^{(3)}$ leads to an accuracy of order $O(n^{-1})$ for "continuous" distributions only. It is surprising that Welch resigns himself to formal expansions even though he proposed already in 1939 an idea which leads easily to a proof (see the remark before Lemma 6). Bartlett (1953) suggested transforming $\sum_{i=1}^n 1^{(1)}(x_i, \theta)$ by some polynomial transformation in such a way that the skewness of the distribution of the transformed variable would become 0. The transformation obtained in this way is nothing else than the Cornish–Fisher transformation, the use of which was suggested by Kendall–Stuart (1961, Section 20.15). Bol'shev (1959, Example 2) provided a justification for this transformation and pointed out that it would be more practical to approximate the $(1-\alpha)$-quantile of the distribution of $\sum_{i=1}^n 1^{(1)}(x_i, \theta)$ rather than to transform this variable. But even then one has to solve $\{\mathbf{x} \in X^{\mathrm{N}} : \sum_{i=1}^n 1^{(1)}(x_i, \theta) \geqq \lambda_{n,\alpha}(\theta)\}$ for θ, which does not necessarily lead to an interval. Hence the confidence interval provided by Theorem 4 has distinctive advantages. Neither Bartlett nor Kendall–Stuart nor Bol'shev mention that this procedure is justified only for "continuous" distributions. The suggestion of Wilks (1963, pages 366—368) to obtain a confidence procedure from $\sum_{i=1}^n F(x_i, \theta)$ (where F is the distribution function) has the advantage of an exact confidence coefficient. It will, however, be approximately optimal only under exceptional circumstances.

6. Estimators. In Theorem 4 we obtained a confidence interval $(F_{n,\alpha}^{(s)}(\theta_n(\mathbf{x})), \infty)$ with confidence coefficient $1 - \alpha + o(n^{-(s-2)/2})$. A corresponding argument leads to the confidence interval $(-\infty, \hat{F}_{n,\alpha}^{(s)}(\theta_n(\mathbf{x}))$ with confidence coefficient $1 - \alpha + o(n^{-(s-2)/2})$. For $\alpha = \frac{1}{2}$ we have $F_{n,\frac{1}{2}}^{(s)}(\theta_n(\mathbf{x})) = \hat{F}_{n,\frac{1}{2}}^{(s)}(\theta_n(\mathbf{x})) \equiv \theta_n^{(s)}(\mathbf{x})$, say. We have

$$(6.1)\qquad \theta_n^{(s)}(\mathbf{x}) = \theta_n(\mathbf{x}) + \sum_{m=1}^{[(s-1)/2]} n^{-m} b_{2m-1,0}^*(\theta_n(\mathbf{x}))$$

(where the functions $b_{m,0}^*$ are defined as in Theorem 4).

$\theta_n^{(s)}$, considered as an estimator for θ, will be median unbiased up to an error term of order $o(n^{-(s-2)/2})$. From Theorem 5 (iii) we expect that $\theta_n^{(s)}$ will be more concentrated about θ (up to an error term of order $o(n^{-\frac{1}{2}})$) than any other estimator which is median unbiased up to an error term of order $o(n^{-\frac{1}{2}})$, provided $\theta_n^{(s)}$ is based on the m.l. estimator θ_n.

DEFINITION. A sequence of estimators T_n, $n \in \mathbb{N}$, is *approximately median unbiased* of order $o(n^{-r})$ if

$$P_\theta^{\mathbb{N}}\{\mathbf{x} \in X^{\mathbb{N}} : T_n(\mathbf{x}) \geqq \theta\} \geqq \tfrac{1}{2} - o(n^{-r}) \tag{6.2}$$

and

$$P_\theta^{\mathbb{N}}\{\mathbf{x} \in X^{\mathbb{N}} : T_n(\mathbf{x}) \leqq \theta\} \geqq \tfrac{1}{2} - o(n^{-r}) . \tag{6.3}$$

THEOREM 6. (i) *Assume that regularity condition* (xii) *and the regularity conditions specified in Theorem* 3 (i) *are fulfilled. If* T_n, $n \in \mathbb{N}$, *is any sequence of estimators which is approximately median unbiased of order* $o(n^{-\frac{1}{2}})$, *uniformly on some compact subset* $K \subset \Theta$, *then uniformly for* $\theta \in K$ *and uniformly for* $t', t'' > 0$:

$$\begin{aligned} P_\theta^{\mathbb{N}}\{\mathbf{x} \in X^{\mathbb{N}} : \theta - n^{-\frac{1}{2}}t' < T_n(\mathbf{x}) < \theta + n^{-\frac{1}{2}}t''\} \\ \leqq \Phi(t''L_2(\theta)^{\frac{1}{2}}) - \Phi(-t'L_2(\theta)^{\frac{1}{2}}) \\ + n^{-\frac{1}{2}}[\varphi(t''L_2(\theta)^{\frac{1}{2}})t''^2 - \varphi(t'L_2(\theta)^{\frac{1}{2}})t'^2] \\ \times L_2(\theta)^{-\frac{1}{2}}[\tfrac{1}{3}L_3(\theta) + \tfrac{1}{2}L_{11}(\theta)] + o(n^{-\frac{1}{2}}) . \end{aligned} \tag{6.4}$$

We remark that a sequence $(T_n)_{n \in \mathbb{N}}$ *fulfilling* (6.4) *is necessarily approximately median unbiased of order* $o(n^{-\frac{1}{2}})$.

(ii) *If the regularity conditions specified in Theorem* 4 *are fulfilled for the likelihood contrast function, then equality holds in* (6.4) *if* $T_n = \theta_n^{(s)}$, *provided the adjusted estimator* $\theta_n^{(s)}$ *is based on the* m.l. *estimator. The equality holds uniformly for* $\theta \in K$ *and* $0 < t', t'' \leqq c_K n^{\frac{1}{2}}$.

(iii) *If the regularity conditions specified in* (i) *and* (ii) *are fulfilled, then the following relation holds uniformly for* $\theta \in K$, $0 < t', t'' \leqq c_K n^{\frac{1}{2}}$, *provided the adjusted estimator* $\theta_n^{(s)}$ *is based on the* m.l. *estimator.*

$$\begin{aligned} P_\theta^{\mathbb{N}}\{\mathbf{x} \in X^{\mathbb{N}} : \theta - n^{-\frac{1}{2}}t' < T_n(\mathbf{x}) < \theta + n^{-\frac{1}{2}}t''\} \\ \leqq P_\theta^{\mathbb{N}}\{\mathbf{x} \in X^{\mathbb{N}} : \theta - n^{-\frac{1}{2}}t' < \theta_n^{(s)}(\mathbf{x}) < \theta + n^{-\frac{1}{2}}t''\} + o(n^{-\frac{1}{2}}) . \end{aligned} \tag{6.5}$$

Under appropriate regularity conditions, (6.5) even holds with $o(n^{-\frac{1}{2}})$ replaced by $O(n^{-1})$, uniformly in θ, t', t'' on compact subsets provided T_n is approximately median unbiased of order $O(n^{-1})$. The corresponding relation with $o(n^{-\frac{1}{2}})$ replaced by $o(n^{-1})$ does not hold true any more in general. For a formulation of this statement in terms of deficiency see the remark following Theorem 3. Theorem 6 (iii) shows that the adjusted m.l. estimator is "second-order efficient" in the class of all estimators which are approximately median unbiased of order $o(n^{-\frac{1}{2}})$. The availability of such an estimator should lead us to reconsider the whole theory of BAN-estimators (which yields first-order efficient estimators only in general). We are, by the way, of the opinion that any definition of second order efficiency should be based on covering probabilities, and that (6.5) provides a natural basis for doing so. An unmotivated concept of second order efficiency like that of C. R. Rao (1963) should be abandoned.

We remark that for families of p-measures fulfilling regularity conditions R_2 only (and no additional conditions like "nonlattice" or "C") relation (6.5) with $o(n^{-\frac{1}{2}})$ replaced by $O(n^{-\frac{1}{2}})$ has been obtained earlier by Pfanzagl (1972) with $O(n^{-\frac{1}{2}})$ depending on t', t'', and by Michel (1972) uniformly in t', $t'' > 0$.

Starting from Theorem 6 (iii) one can show in the usual way (see, for instance, Pfanzagl (1970), page 35, Corollary 1.15) that among all approximately median unbiased estimators the adjusted m.l. estimator approximately minimizes the average loss for any bounded loss function $L(\theta, \tau)$ which is non-decreasing as τ moves away from θ in either direction.

Recall that an exponential family has monotone likelihood ratios for arbitrary sample sizes and therefore (see Lehmann (1959), page 83, for continuous and Pfanzagl (1970), page 33, Theorem 1.12, for arbitrary distributions) admits a strictly median unbiased (randomized) estimator which is maximally concentrated in the class of all median unbiased estimators. Because of the use of randomized estimators, this result also holds true for discrete distributions.

Theorem 6 strongly suggests the use of

$$\theta_n^*(\mathbf{x}) \equiv \theta_n(\mathbf{x}) + \frac{n^{-1}}{6} L_2(\theta_n(\mathbf{x}))^{-2} L_3(\theta_n(\mathbf{x})) \tag{6.6}$$

instead of the m.l. estimator θ_n. (Notice that $\theta_n^{(3)} = \theta_n^{(4)} = \theta_n^*$.)

As an estimator, θ_n^* (or any other $\theta_n^{(s)}$ for $s \geqq 3$) is superior to θ_n because of its smaller median bias. As a base for the construction of critical regions or confidence procedures it is equivalent to θ_n (see the following Remark). It shares in particular the optimum properties of θ_n as formulated in Theorems 3 (iii) and 5 (iii).

REMARK. Assume that the regularity conditions R_s are fulfilled for some $s \geqq 3$. Assume that for $\nu = 1, \cdots, [(s-1)/2]$ the function $b^*_{2\nu-1,0}$ is $(s-2\nu)$-times differentiable, and that the derivative of order $(s-2\nu)$ fulfills a Lipschitz-condition. Then the following holds true:

(i) For any sequence of critical regions (4.2) based on θ_n there exists a sequence of critical regions based on $\theta_n^{(s)}$ which has the same power function up to terms of order $o(n^{-(s-2)/2})$.

(ii) Uniformly on compact subsets of Θ and uniformly for $t \in \mathbb{R}$:

$$P_\theta^{\mathbb{N}}\left\{\mathbf{x} \in X^{\mathbb{N}} : n^{\frac{1}{2}} \frac{\theta_n^{(s)}(\mathbf{x}) - \theta}{\beta(\theta)} \leqq t\right\}$$
$$= \Phi(t) + \varphi(t) \textstyle\sum_{m=1}^{s-2} n^{-m/2} Q_m(t, \theta) + o(n^{-(s-2)/2}) .$$

We have in particular

$$Q_1(t, \cdot) = q_{11} t^2 , \qquad Q_2(t, \cdot) = q_{20} t + q_{21} t^3 + q_{22} t^5$$

with

$$q_{11} = a_{11}$$
$$q_{20} = \tfrac{1}{2}a_{10}^2 - 2a_{10}a_{11} + a_{20} - (a_{10}\beta)'$$
$$q_{21} = a_{21}$$
$$q_{22} = a_{22}$$

(where the a_{ij} are defined as in Theorem 1).

7. Some examples. To illustrate the general results we shall apply them to the particular case of location-parameters and scale-parameters. The more relevant problem of estimating location- and scale-parameters simultaneously requires corresponding results for vector parameters which are not yet available.

(a) *Location parameters.* Let $p: \mathbb{R} \to (0, \infty)$ be a continuous function with $\int p(x)\,dx = 1$. For $\theta \in \mathbb{R}$ let P_θ denote the p-measure with Lebesgue density $x \to p(x - \theta)$.

In this case the quantities $L_{ijk}(\theta)$ are independent of θ so that Theorem 1 can be immediately applied to obtain critical regions.

The approximately median unbiased estimator is

$$\theta_n^*(x_1, \cdots, x_n) = \theta_n(x_1, \cdots, x_n) + \frac{n^{-1}}{6} L_2^{-2} L_3 .$$

If p is symmetric about 0, we have $L_3 = 0$ so that the m.l. estimator is median unbiased up to an error of order $o(n^{-1})$ $[O(n^{-\frac{3}{2}})]$.

The following conditions can be easily verified in the most common cases. They imply the regularity conditions R_s for the likelihood contrast function

$$f(x, \theta) \equiv -1(x - \theta) \qquad \text{for} \quad \theta \in \mathbb{R}$$
$$\equiv +\infty \qquad \text{for} \quad \theta = \pm\infty .$$

(7.1) $\quad p$ is unimodal,

(7.2) $\quad \int |1(x)|^{s-1} p(x + t)\,dx < \infty \quad$ for some $t > 0$ and some $t < 0$,

(7.3) $\quad 1^{(r)}$ is bounded for $r = 1, \cdots, s + 1$.

The proof is more or less straightforward and will therefore be omitted.

Conditions (7.1)—(7.3) are, of course, fulfilled for the location-parameter family of normal distributions, but our results are irrelevant in this case since the m.l. estimator is exactly normally distributed. Other examples of common location-parameter families are the Cauchy distribution and the logistic distribution. In both cases the verification of (7.1) and (7.3) for any $s \geqq 3$ is straightforward. The verification of (7.2) is slightly more complicated but without substantial difficulties.

The validity of Condition C_* follows from the Criterion in Section 4 or by direct application of the Riemann–Lebesgue theorem. Hence all our results (including the optimum assertion for the m.l. estimator) apply to the location parameter family of Cauchy distributions as well as to the location parameter family of logistic distributions.

(b) *Scale parameters.* Let $p: X \to (0, \infty)$ with $X = \mathbb{R}$ or $X = (0, \infty)$ be a continuous function with $\int_X p(x)\,dx = 1$. For $\theta \in (0, \infty)$ let P_θ denote the p-measure with Lebesgue density $x \to \theta^{-1}p(\theta^{-1}x)$. Though the L_{ijk} depend on θ in this case, the a_{ij} are independent of θ so that, again, Theorem 1 could be applied immediately to obtain critical regions.

The approximately median unbiased estimator is

$$\theta_n^*(x_1, \cdots, x_n) = \theta_n(x_1, \cdots, x_n)\left(1 + \frac{n^{-1}}{6} L_2(1)^{-2}L_3(1))\right).$$

The verification of regularity conditions R_s can be simplified by application of the following relation:

Let p be unimodal. (Then p is either bounded or $X = (0, \infty)$ and p has its mode at 0.) Let $g: \mathbb{R} \to [0, \infty]$ be measurable. If $\int g(x)p(tx)\,dx < \infty$ for some $t \in (0, 1)$, then for every $\theta > 0$ there exist $\theta' < \theta < \theta''$ such that

$$\sup_{\tau \in (\theta', \theta'')} \int g(\theta x)p(\tau x)\tau\,dx < \infty .$$

This relation is useful for the proofs of conditions (iii) and (vii).

An elementary computation shows that in the case of the scale parameter family of the exponential, normal, Cauchy and logistic distribution, the regularity conditions R_s are fulfilled for all $s \geqq 3$. The verification of Condition C_* is slightly more complex.

To illustrate the accuracy of the results we shall study the exponential distribution

$$p(x) = \exp[-x] \qquad \text{for} \quad x > 0$$

in more detail, since in this case the exact distribution of the m.l. estimator is easily accessible.

First of all we remark that C_* is fulfilled for $h(\cdot, \theta, \tau) = 1^{(1)}(\cdot, \tau)$, $\tau \in \bar{\Theta}$, since $P_\theta * 1^{(1)}(\cdot, \tau)$ has relative to the Lebesgue measure the density

$$\begin{aligned} q_\theta(r, \tau) &\equiv \tau^2\theta^{-1}\exp[-\tau^2\theta^{-1}(r + \tau^{-1})] \qquad &\text{if} \quad r > -\tau^{-1}, \\ &\equiv 0 &\text{if} \quad r \leqq -\tau^{-1} \end{aligned}$$

which is continuous in τ at $\tau = \theta$ for all $r \neq \theta^{-1}$.

Condition C_* is also fulfilled for

$$\begin{aligned} h(\cdot, \theta, \tau) &\equiv (\theta - \tau)^{-1}(1(\cdot, \theta) - 1(\cdot, \tau)) \qquad & \theta \neq \tau, \\ &\equiv 1^{(1)}(\cdot, \tau) & \theta = \tau \quad \theta, \tau \in \bar{\Theta} \end{aligned}$$

but this is irrelevant in this case since the family has monotone likelihood ratios so that the critical region (7.4) is uniformly most powerful (see Lehmann, page 68, Theorem 2) and the median unbiased estimator derived from the m.l. estimator is uniformly most concentrated (see Lehmann, page 83 or Pfanzagl (1970), page 33, Theorem 1.12), so that Theorems 3 and 6 are irrelevant.

From Theorem 2 we obtain the critical region

$$\left\{(x_1, \cdots, x_n) \in \mathbb{R}^n : \frac{1}{n} \sum_1^n x_i \geqq G_{n,\alpha}^{(s)}(\theta)\right\} \tag{7.4}$$

where

$$G_{n,\alpha}^{(2)}(\theta) = \theta(1 - n^{-\frac{1}{2}}N_\alpha)$$

is the usual normal approximation, and

$$G_{n,\alpha}^{(3)}(\theta) = \theta\left(1 - n^{-\frac{1}{2}}N_\alpha + \frac{n^{-1}}{3}(N_\alpha^2 - 1)\right).$$

$$G_{n,\alpha}^{(4)}(\theta) = \theta\left(1 - n^{-\frac{1}{2}}N_\alpha + \frac{n^{-1}}{3}(N_\alpha^2 - 1) + \frac{n^{-\frac{3}{2}}}{36}N_\alpha(7 - N_\alpha^2)\right).$$

The following table gives the actual error of type one which is achieved by this approximation formula: The entry in the columns is $P_\theta^n\{(x_1, \cdots, x_n) \in \mathbb{R}^n : (1/n) \sum_1^n x_i \geqq G_{n,\alpha}^{(s)}(\theta)\}$ for $\alpha = .05$ and varying sample sizes n.

TABLE 1
Actual error of type one for $\alpha = .05$

Sample size n	$s = 2$	$s = 3$	$s = 4$
5	.0668	.0472	.0498
10	.0636	.0485	.0499
25	.0596	.0493	.0500
50	.0572	.0496	.0500
100	.0553	.0498	.0500

The result of this comparison can be summarized in one sentence: Even with one additional term to the normal approximation ($s = 3$) the accuracy for the sample size $n = 5$ is greater than the accuracy of the normal approximation for the sample size $n = 100$.

The results concerning confidence intervals are less satisfactory. From Theorem 4 we obtain the confidence interval

$$\left(F_{n,\alpha}^{(s)}\left(\frac{1}{n} \sum_{i=1}^n x_i\right), \infty\right) \tag{7.5}$$

where

$$F_{n,\alpha}^{(2)}(\theta) = \theta(1 + n^{-\frac{1}{2}}N_\alpha)$$
$$F_{n,\alpha}^{(3)}(\theta) = \theta(1 + n^{-\frac{1}{2}}N_\alpha + n^{-1}\tfrac{1}{3}(1 + 2N_\alpha^2))$$
$$F_{n,\alpha}^{(4)}(\theta) = \theta(1 + n^{-\frac{1}{2}}N_\alpha + n^{-1}\tfrac{1}{3}(1 + 2N_\alpha^2) + n^{-\frac{3}{2}}\tfrac{1}{36}(17N_\alpha + 13N_\alpha^3)).$$

In order to facilitate the comparison with Table 1, the following Table 2 contains not the confidence coefficient, but its complementary value, i.e. the probability of not covering the parameter value. The entry in the columns is $P_\theta^n\{(x_1, \cdots, x_n) \in \mathbb{R}^n : \theta > F_{n,\alpha}^{(s)}((1/n) \sum_{i=1}^n x_i)\}$ for $\alpha = .05$ and varying sample size n.

TABLE 2
Probability of not covering the true parameter value for $\alpha = .05$

Sample size n	$s = 2$	$s = 3$	$s = 4$
5	.0000	.1532	.0218
10	.0030	.0910	.0397
25	.0139	.0632	.0475
50	.0225	.0560	.0492
100	.0297	.0528	.0497

This table shows that for small sample sizes the accuracy is insufficient for $s < 4$. Since $G_{n,\alpha}^{(s)}(\theta) = \theta G_{n,\alpha}^{(s)}(1)$ in our particular example, one could, of course, also use the confidence intervals

$$\left((G_{n,\alpha}^{(s)}(1))^{-1}\left(\frac{1}{n}\sum_1^n x_i\right), \infty\right),$$

the confidence coefficient of which is in better agreement with the prescribed value $1 - \alpha$ than that of the confidence intervals (7.5). We have abstained from doing so since our aim is to check the accuracy of Theorem 4 and not to obtain good confidence intervals for the exponential distribution.

By (6.6), the adjusted m.l. estimator is

$$\theta_n^*(x_1, \cdots, x_n) = \left(1 + \frac{1}{3n}\right)\frac{1}{n}\sum_1^n x_i .$$

The following Table 3 shows in the column θ_n the probability $P_\theta^n\{(x_1, \cdots, x_n) \in R^n : (1/n)\sum_1^n x_i \geqq \theta\}$, in the column θ_n^* the probability $P_\theta^n\{(x_1, \cdots, x_n) \in \mathbb{R}^n : (1 + (1/3n))(1/n)\sum_1^n x_i \geqq \theta\}$.

TABLE 3
Median bias

Sample size n	θ_n	θ_n^*
5	.441	.497
10	.458	.499
25	.473	.500
50	.481	.500
100	.487	.500

8. Assumptions. This section collects regularity conditions which are needed for the proofs, as well as conditions related to Cramér's Condition C.

(i) $\theta \to P_\theta$ is continuous in Θ with respect to the supremum-metric on $\{P_\theta : \theta \in \Theta\}$ (defined by the distance function

$$d(P, Q) = \sup_{A \in \mathscr{A}} |P(A) - Q(A)|) .$$

(ii) For each $x \in X$, $\theta \to f(x, \theta)$ is continuous on $\bar{\Theta}$

(iii) For every $\theta \in \Theta$ there exists a neighborhood U_θ such that

$$\sup_{\tau \in U_\theta} E_\tau(|f(\cdot, \theta)|^{s-1}) < \infty .$$

(iv) For each $x \in X$, $\theta \to f(x, \theta)$ is twice differentiable in Θ for all $\theta \in \Theta$ and

$$E_\theta(f^{(1)}(\cdot, \theta)) = 0 .$$

(v) For every $\theta \in \Theta$ there exists a neighborhood U_θ such that

(a) $\inf_{\tau \in U_\theta} E_\tau(f^{(1)}(\cdot, \tau)^2) > 0$,

(b) $\inf_{\tau \in U_\theta} E_\tau(f^{(2)}(\cdot, \tau)) > 0$.

(vi) For every $\theta \in \bar{\Theta}$, $\tau \in \Theta$ there exist neighborhoods V_θ, W_τ such that for all neighborhoods $V \subset V_\theta$ of θ, $\sup_{\sigma \in W_\tau} E_\sigma(|\inf_{\delta \in V} f(\cdot, \delta)|^{s-1}) < \infty$.

(vii) $f^{(2)}(\cdot, \theta)$, $\theta \in \Theta$ fulfills Condition L_{s-1}.

(viii) For every $\theta \in \Theta$ there exists a neighborhood U_θ such that

$$\lim_{A \to \infty} \sup_{\tau \in U_\theta} \sup_{\delta \in U_\theta} E_\tau(|f^{(1)}(\cdot, \delta)|^s 1_{\{x \in X : |f^{(1)}(x, \delta)| > A\}}) = 0 .$$

(ix) For each $x \in X$, $\theta \to f(x, \theta)$ is $(s + 1)$-times differentiable in Θ, and $\theta \to f^{(s+1)}(x, \theta)$ is continuous.

(x) For every $\theta \in \Theta$ there exists a neighborhood U_θ such that for all $k_\nu \in \mathbb{N} \cup \{0\}$, $\nu = 1, \cdots, s + 1$ with $k_\nu < s$ and $\sum_{\nu=1}^{s+1} \nu k_\nu \leqq s + 1$:

$$\sup_{\tau \in U_\theta} \sup_{\delta \in U_\theta} E_\tau(\prod_{\nu=1}^{s+1} (f^{(\nu)}(\cdot, \delta)^{k_\nu})) < \infty .$$

(xi) For every $\theta \in \Theta$ there exists a neighborhood U_θ and a function $m(\cdot, \theta)$ such that for all $x \in X$, δ, $\tau \in U_\theta$

$$\left| \frac{p(x, \delta)}{p(x, \tau)} - 1 \right| \leqq |\delta - \tau| m(x, \theta)$$

and

$$\sup_{\tau \in U_\theta} E_\tau(m(\cdot, \tau)^4) < \infty .$$

(We remark that condition (xi) implies condition (viii) with $s = 4$ for the likelihood contrast function.)

(xii) L_2 is twice differentiable in Θ, L_3 and L_{11} are differentiable in Θ, and for every $\theta \in \Theta$ there exists a neighborhood U_θ such that

(a) $\sup_{\tau \in U_\theta} |L_2''(\tau)| < \infty$

(b) $\sup_{\tau \in U_\theta} |L_3'(\tau)| < \infty$

(c) $\sup_{\tau \in U_\theta} |L_{11}'(\tau)| < \infty$.

For notational convenience, conditions (i)—(x) will be labeled as *Condition* R_s.

CONDITION L_r. A family of functions $g(\cdot, \theta)$, $\theta \in \Theta$, fulfills Condition L_r if for every $\theta \in \Theta$ there exists a neighborhood U_θ and a function $m(\cdot, \theta)$ such that

(a) for all $x \in X$, δ, $\tau \in U_\theta$, $|g(x, \tau) - g(x, \delta)| \leqq |\tau - \delta| m(x, \theta)$

(b) $\sup_{\tau \in U_\theta} E_\tau(m(\cdot, \theta)^r) < \infty$

(c) $\sup_{\tau \in U_\theta} E_\tau(|g(\cdot, \tau)|^r) > \infty$.

In addition to the regularity conditions specified above we have to assume that the family $f^{(1)}(\cdot, \theta)$, $\theta \in \Theta$, fulfills a certain continuity condition. For the optimality assertions (Theorem 3) we need a related continuity condition for $1(\cdot, \theta)$, $\theta \in \Theta$. The formulation of this continuity condition will be based on the following definition.

CONDITION C_*. A family of $\mathscr{A}$-measurable functions $h(\cdot, \theta, \tau): X \to \overline{\mathbb{R}}$ θ, $\tau \in \Theta$ fulfills Condition C_* if for every $\theta \in \Theta$ there exists an open neighborhood U_θ of θ such that

$$(*) \qquad \limsup_{|u|\to\infty} \sup_{\delta,\tau \in U_\theta} |\int \exp[iuh(x, \delta, \tau)]P_\delta(dx)| < 1 .$$

The following criterion makes the nature of Condition C_* more transparent. It is also useful for verifying this condition.

CRITERION. Assume that $P_\theta | \mathscr{A}$, $\theta \in \Theta$, admits densities which are continuous functions of θ. If the family of induced measures $P_\theta * h(\cdot, \delta, \tau) | \mathscr{B}$, $\delta, \theta, \tau \in \Theta$, admits densities relative to the Lebesgue measure, say $q_\theta(\cdot, \delta, \tau)$, such that for every $\theta \in \Theta$ and every $r \in \mathbb{R}$ (with the exception of r belonging to a Lebesgue null set possibly depending on θ) the map $(\delta, \tau) \to q_\theta(r, \delta, \tau)$ is continuous at $(\delta, \tau) = (\theta, \theta)$, then $h(\cdot, \theta, \tau)$, θ, $\tau \in \Theta$, fulfills Condition C_*.

PROOF. Using Scheffé's lemma (see e.g. Lehmann, page 351, Lemma 4) we obtain the existence of a neighborhood U_θ of θ such that for all δ, $\tau \in U_\theta$

$$\int |q_\delta(r, \delta, \tau) - q_\theta(r, \delta, \tau)|\, dr = 2 \sup_{B \in \mathscr{B}} |P_\delta * h(\cdot, \delta, \tau)(B) - P_\theta * h(\cdot, \delta, \tau)(B)|$$
$$= 2 \sup_{A \in \mathscr{A}} |P_\delta(A) - P_\theta(A)| \leqq \tfrac{1}{4} ,$$

and for all δ, $\tau \in U_\theta$

$$\int |q_\theta(r, \delta, \tau) - q_\theta(r, \theta, \theta)|\, dr \leqq \tfrac{1}{4} .$$

Hence δ, $\tau \in U_\theta$ implies

$$(8.1) \qquad \int |q_\delta(r, \delta, \tau) - q_\theta(r, \theta, \theta)|\, dr \leqq \tfrac{1}{2} .$$

By the Riemann-Lebesgue theorem (see e.g. Feller **2**, page 486, Lemma 3) for every $\theta \in \Theta$ there exists $k_\theta > 0$ such that for $|u| \geqq k_\theta$

$$|\int \exp[iur]q_\theta(r, \theta, \theta)\, dr| \leqq \tfrac{1}{4} .$$

Hence $|u| \geqq k_\theta$ and δ, $\tau \in U_\theta$ imply by (8.1):

$$\begin{aligned} &|\int \exp[iuh(x, \delta, \tau)]P_\delta(dx)| \\ &\qquad = |\int \exp[iur]q_\delta(r, \delta, \tau)\, dr| \\ &\qquad \leqq |\int \exp[iur]q_\delta(r, \delta, \tau)\, dr - \int \exp[iur]q_\theta(r, \theta, \theta)\, dr| \\ &\qquad\quad + |\int \exp[iur]q_\theta(r, \theta, \theta)\, dr| \leqq \tfrac{3}{4} . \end{aligned}$$

9. Lemmas.

LEMMA 1. *Let $g(\cdot, \theta)$, $\theta \in \Theta$, be a family of $\mathscr{A}$-measurable functions fulfilling*

(i) $E_\theta(g(\cdot, \theta)) = 0$ *and* (ii) $E_\theta(g(\cdot, \theta)^2) = 1$ *for all* $\theta \in \Theta$.

Assume, furthermore, that for some compact $K \subset \Theta$ and some $s \geqq 3$:

(iii) $\sup_{\theta \in K} E_\theta(|g(\cdot, \theta)|^s) < \infty$.

Then uniformly for $\theta \in K$:

$$P_\theta^{\mathrm{N}}\{\mathbf{x} \in X^{\mathrm{N}} : |n^{-\frac{1}{2}} \sum_{i=1}^{n} g(x_i, \theta)| > (2(s-2)\log n)^{\frac{1}{2}}\} = o(n^{-(s-2)/2}) .$$

PROOF. Follows from Esséen, page 73, Theorem 2 and Feller **1**, page 166, Lemma 2 or from Michel (1972) with $(2(s-2)\log n)^{\frac{1}{2}}$ replaced by $((s-2)\log n)^{\frac{1}{2}}$ (see the remark following Theorem 4 in Michel).

LEMMA 2. *Assume that* $g(\cdot, \theta)$, $\theta \in \Theta$, *fulfills the conditions* (i), (iii) *specified in Lemma* 1 *for some* $s \geqq 2$. *Then for every* $u > 0$:

$$P_\theta^{\mathrm{N}}\{\mathbf{x} \in X^{\mathrm{N}} : |n^{-\frac{1}{2}} \sum_{i=1}^{n} g(x_i, \theta)| > u\} \leqq c_K u^{-s} .$$

PROOF. Let $g_n(\mathbf{x}, \theta) \equiv n^{-\frac{1}{2}} \sum_{i=1}^{n} g(x_i, \theta)$. We have

$$P_\theta^{\mathrm{N}}\{\mathbf{x} \in X^{\mathrm{N}} : |g_n(\mathbf{x}, \theta)| > u\} \leqq u^{-s} E_\theta^{\mathrm{N}}(|g_n(\cdot, \theta)|^s)$$

and $E_\theta^{\mathrm{N}}(|g_n(\cdot, \theta)|^s) \leqq C(s) E_\theta(|g(\cdot, \theta)|^s)$, by inequality (3.3) in Chung (1951), page 341.

LEMMA 3. *Assume that regularity conditions* (i)—(viii) *are fulfilled for some* $s \geqq 3$. *Then uniformly on compact subsets of* Θ:

$$P_\theta^{\mathrm{N}}\{\mathbf{x} \in X^{\mathrm{N}} : \frac{|\theta_n(\mathbf{x}) - \theta|}{\beta(\theta)} n^{\frac{1}{2}} \geqq (4(s-2)\log n)^{\frac{1}{2}}\} = o(n^{-(s-2)/2}) .$$

For $s = 3$ this lemma reduces to Lemma 6 in Michel and Pfanzagl (1971), page 82. The proof will be omitted since it is almost the same, the only exception being the use of Lemma 1 instead of the Berry–Esséen theorem.

Lemmas 4 and 5 of Michel and Pfanzagl, which are needed for the proof of Lemma 6, can be obtained with cn^{-1} replaced by $cn^{-(s-1)/2}$ if Lemma 2 (for $s-1$) is used instead of Chebychev's inequality.

LEMMA 4. *Let* $s \geqq 3$. *Assume that* $g(\cdot, \theta, \tau)$, $\theta, \tau \in \Theta$, *is a family of* $\mathscr{A}$*-measurable functions fulfilling Condition* C_* *and the conditions*

(i) $E_\theta(g(\cdot, \theta, \tau)) = 0$, $E_\theta(g(\cdot, \theta, \tau)^2) = 1$ *for all* $\theta, \tau \in \Theta$,
(ii) $\sup_{\theta \in K} \sup_{|\tau-\theta|<e_K} E_\theta(|g(\cdot, \theta, \tau)|^s) < \infty$ *for some compact* $K \subset \Theta$, $e_K > 0$.

Then there exists $a_K > 0$ *such that uniformly for* $\theta \in K$, $|\tau - \theta| < a_K$, *and* $u \in \mathbb{R}$

$$P_\theta^{\mathrm{N}}\{\mathbf{x} \in X^{\mathrm{N}} : n^{-\frac{1}{2}} \sum_{i=1}^{n} g(x_i, \theta, \tau) < u\} = \Phi(u) + \varphi(u) \sum_{m=1}^{s-2} n^{-m/2} Q_m(u, \theta, \tau) + o(n^{-(s-2)/2}) ,$$

where $Q_m(u, \theta, \tau)$ *is a polynomial in* u *with coefficients depending on the moments* $\alpha_k \equiv E_\theta(g(\cdot, \theta, \tau)^k)$, $k = 3, \cdots, m+2$.

We have in particular

$$Q_1(u, \cdot, \cdot) = q_{10} + q_{11}u^2$$
$$Q_2(u, \cdot, \cdot) = q_{20}u + q_{21}u^3 + q_{22}u^5$$

with

$$q_{10} = \tfrac{1}{6}\alpha_3$$
$$q_{11} = -\tfrac{1}{6}\alpha_3$$
$$q_{20} = \tfrac{1}{24}(3\alpha_4 - 5\alpha_3^2 - 9)$$
$$q_{21} = \tfrac{1}{72}(-3\alpha_4 + 10\alpha_3^2 + 9)$$
$$q_{22} = -\tfrac{1}{72}\alpha_3^2 .$$

This lemma is nothing else than a uniform version of the well-known Edgeworth-expansion (see e.g. Gnedenko-Kolmogorov page 220, Theorem).[2] Since our Condition C_* is a uniform version of Cramér's Condition C, the proof goes through without substantial changes. We need only remark that C_* implies that for every compact $K \subset \Theta$ there exists $a_K > 0$ such that

$$\limsup_{|u|\to\infty} \sup_{\delta\in K} \sup_{|\tau-\delta|<a_K} |\textstyle\int \exp[iuh(x, \delta, \tau)]P_\delta(dx)| < 1 .$$

This can easily be seen as follows: for every $\theta \in \Theta$ let $a_\theta \in (0, e_K)$ be such that $(\theta - 2a_\theta, \theta + 2a_\theta) \subset U_\theta$. Since $\{(\theta - a_\theta, \theta + a_\theta): \theta \in K\}$ is an open cover of K, there exists a finite subcover determined by the values $\theta_1, \cdots, \theta_r$. Then the assertion holds true with $a_K \equiv \min\{a_{\theta_1}, \cdots, a_{\theta_r}\}$ (since $\delta \in (\theta_i - \alpha_{\theta_i}, \theta_i + a_{\theta_i})$ together with $|\tau - \delta| < a_K \leqq a_{\theta_i}$ implies $\delta, \tau \in U_{\theta_i}$).

LEMMA 5. *Assume that $g(\cdot, \theta)$, $\theta \in \Theta$, fulfills Condition L_r for $r \geqq 2$. Then for every compact $K \subset \Theta$ there exist constants $b_K > 0$, $d_K > 0$, and for every $\theta \in K$, $n \in \mathbb{N}$ a set $A_{n,\theta,K} \in \mathscr{A}^n$ such that*

(a) $E_\theta(\sup_{|\tau-\theta|\leqq d_K} |g(\cdot, \tau)|^r) \leqq b_K$,
(b) $\sup_{\theta\in K} P_\theta^{\mathbb{N}}(X^{\mathbb{N}} - A_{n,\theta,K}) = O(n^{-r/2})$,
(c) $\mathbf{x} \in A_{n,\theta,K}$, $\tau \in \theta$ *and* $|\tau - \theta| \leqq d_K$ *imply*

$$|n^{-1}\textstyle\sum_{i=1}^n g(x_i, \tau) - n^{-1}\sum_{i=1}^n g(x_i, \theta)| \leqq |\tau - \theta| b_K .$$

The proof is standard and will therefore be omitted.

Lemma 6 makes precise the following idea: Since

$$O = \textstyle\sum_{i=1}^n f^{(1)}(x_i, \theta_n(\mathbf{x})) = \sum_{i=1}^n f^{(1)}(x_i, \tau) + (\theta_n(x) - \tau)\sum_{i=1}^n f^{(2)}(x_i, \theta_n(\mathbf{x}, \tau))$$

and $\sum_{i=1}^n f^{(2)}(x_i, \theta_n(\mathbf{x}, \tau)) > 0$ with high probability, we have with high probability:

$$\textstyle\sum_{i=1}^n f^{(1)}(x_i, \tau) > 0 \quad \text{iff} \quad \theta_n(\mathbf{x}) < \tau .$$

This idea occurs already in Welch (1939, page 188), albeit the conclusion $P_\theta^{\mathbb{N}}\{\mathbf{x} \in X^{\mathbb{N}}: \sum_{i=1}^n f^{(1)}(x_i, \tau) > 0\} = P_\theta^{\mathbb{N}}\{\mathbf{x} \in X^{\mathbb{N}}: \theta_n(\mathbf{x}) < \tau\}$ which he draws from this idea (see Welch 1965, page 6, formula (42)) is not correct.

LEMMA 6. *Let θ_n, $n \in \mathbb{N}$, be a sequence of* m.c. *estimators. Assume that for some $s \geqq 3$ regularity conditions* (i)—(viii) *are fulfilled.*

[2] The reader who wants to check with Gnedenko-Kolmogorov whether the assertion of Lemma 4 is correct should not be confused by the fact that our expression for Q_2 is not the same as that given by Gnedenko-Kolmogorov, page 195, formula (24). The expression for Q_2 given there has the wrong sign (in all editions, by the way).

Then for every compact $K \subset \Theta$ there exists a constant $a_K > 0$ such that uniformly on $\{(\theta, \tau) \in \Theta^2 : \theta \in K, |\tau - \theta| \leqq a_K\}$:

$$P_\theta^{\mathbb{N}}(\{\mathbf{x} \in X^{\mathbb{N}} : \sum_{i=1}^n f^{(1)}(x_i, \tau) > 0\} \Delta \{\mathbf{x} \in X^{\mathbb{N}} : \theta_n(\mathbf{x}) < \tau\}) = o(n^{-(s-2)/2}) .$$

PROOF. (i) At first we shall prove the existence of a constant $a_K > 0$ such that uniformly on K

$$P_\theta^{\mathbb{N}*}(X^{\mathbb{N}} - B_{n,\theta,K}) = o(n^{-(s-2)/2}) , \tag{9.1}$$

where

$$B_{n,\theta,K} \equiv \{\mathbf{x} \in X^{\mathbb{N}} : \sum_{i=1}^n f^{(2)}(x_i, \tau) > 0 \quad \text{for all } \tau \in \Theta \text{ with } |\tau - \theta| \leqq 2a_K\} .$$

(Since $B_{n,\theta,K}$ is not necessarily measurable, we have to use the outer measure $P_\theta^{\mathbb{N}*}$ pertaining to $P_\theta^{\mathbb{N}}$.)

Let b_K, d_K, and $A_{n,\theta,K}$ be given by Lemma 5, applied for $g = f^{(2)}$. Let $\alpha_K \equiv \inf\{E_\theta(f^{(2)}(\cdot, \theta)) : \theta \in K\}$, $a_K \equiv \min\{\frac{1}{2}d_K, \frac{1}{4}\alpha_K, b_K^{-1}\}$.

We have for all $\theta \in K$, $n \in \mathbb{N}$, $\mathbf{x} \in A_{n,\theta,K}$, and all $\tau \in \Theta$ with $|\tau - \theta| \leqq 2a_K$:

$$\begin{aligned} n^{-1} \sum_{i=1}^n f^{(2)}(x_i, \tau) &\geqq n^{-1} \sum_{i=1}^n f^{(2)}(x_i, \theta) - |\tau - \theta| b_K \\ &\geqq n^{-1} \sum_{i=1}^n f^{(2)}(x_i, \theta) - \tfrac{1}{2}\alpha_K . \end{aligned} \tag{9.2}$$

Let

$$C_{n,\theta,K} \equiv \{\mathbf{x} \in X^{\mathbb{N}} : n^{-1} \sum_{i=1}^n f^{(2)}(x, \theta) > \tfrac{1}{2}\alpha_K\} . \tag{9.3}$$

(9.2) implies for all $\theta \in K$, $n \in \mathbb{N}$:

$$A_{n,\theta,K} \cap C_{n,\theta,K} \subset B_{n,\theta,K} . \tag{9.4}$$

Since $P_\theta^{\mathbb{N}}(X^{\mathbb{N}} - C_{n,\theta,K}) = o(n^{-(s-2)/2})$ by Lemma 2 and $P^{\mathbb{N}}(X^{\mathbb{N}} - A_{n,\theta,K}) = o(n^{-(s-2)2})$ by Lemma 5, relation (9.1) follows immediately from (9.4).

(ii) By a Taylor expansion of $\delta \to \sum_{i=1}^n f^{(1)}(x_i, \delta)$ about $\delta = \tau$ we obtain for all $\mathbf{x} \in X^{\mathbb{N}}$ with $\theta_n(\mathbf{x}) \in \Theta$:

$$\begin{aligned} O &= \sum_{i=1}^n f^{(1)}(x_i, \theta_n(\mathbf{x})) \\ &= \sum_{i=1}^n f^{(1)}(x_i, \tau) + (\theta_n(\mathbf{x}) - \tau) \sum_{i=1}^n f^{(2)}(x_i, \theta_n(\mathbf{x}, \tau)) \end{aligned} \tag{9.5}$$

with $\theta_n(\mathbf{x}, \tau)$ somewhere between τ and $\theta_n(\mathbf{x})$.

Let

$$E_{n,\theta,K} \equiv \{\mathbf{x} \in X^{\mathbb{N}} : |\theta_n(\mathbf{x}) - \theta| \leqq a_K\} .$$

By Lemma 3 we have uniformly for $\theta \in K$:

$$P_\theta^{\mathbb{N}}(X^{\mathbb{N}} - E_{n,\theta,K}) = o(n^{-(s-2)/2}) . \tag{9.6}$$

If $|\tau - \theta| \leqq a_K$ and $\mathbf{x} \in E_{n,\theta,K}$ we have $|\theta_n(\mathbf{x}, \theta) - \theta| \leqq 2a_K$. Hence $|\tau - \theta| \leqq a_K$ and $\mathbf{x} \in E_{n,\theta,K} \cap B_{n,\theta,K}$ imply $\sum_{i=1}^n f^{(2)}(x_i, \theta_n(\mathbf{x}, \tau)) > 0$ and therefore by (9.5),

$$\text{“} \sum_{i=1}^n f^{(1)}(x_i, \tau) > 0 \quad \text{iff} \quad \theta_n(\mathbf{x}) < \tau \text{”}.$$

Hence $|\tau - \theta| \leqq a_K$ implies

$$\begin{aligned} \{x \in X^{\mathbb{N}} : \sum_{i=1}^n f^{(1)}(x_i, \tau) > 0\} &\Delta \{x \in X^{\mathbb{N}} : \theta_n(\mathbf{x}) < \tau\} \\ &\subset (X^{\mathbb{N}} - B_{n,\theta,K}) \cup (X^{\mathbb{N}} - E_{n,\theta,K}) . \end{aligned} \tag{9.7}$$

(9.7), (9.1), and (9.6) together imply the assertion.

Proofs of the following Lemmas 7 and 8 will be omitted, since the techniques are standard in the theory of asymptotic expansion. Proofs of similar results can, for instance, be found in Wasow (1956).

LEMMA 7. *Let F_n be a distribution function such that uniformly for $t \in \mathbb{R}$*

$$F_n(t) = \Phi(t) + \varphi(t) \textstyle\sum_{m=1}^{r} n^{-m/2} Q_m(t) + o(n^{-r/2}) . \tag{9.8}$$

(i) *Existence: Then there exist polynomials in t, say $Q_1^*, \cdots, Q_r^*$, the coefficients of Q_m^* being rational functions of the coefficients of $Q_1, \cdots, Q_m$, such that uniformly for $|t| \leqq \log n$*

$$F_n(t + \textstyle\sum_{m=1}^{r} n^{-m/2} Q_m^*(t)) = \Phi(t) + o(n^{-r/2}) . \tag{9.9}$$

We have in particular:

$$Q_1^*(t) = -Q_1(t)$$

$$Q_2^*(t) = Q_1(t) \frac{d}{dt} Q_1(t) - \frac{t}{2} Q_1(t)^2 - Q_2(t) .$$

(ii) *Uniqueness: If $(t_n)_{n \in \mathbb{N}}$ is a sequence such that $F_n(t_n) = \Phi(t_0) + o(n^{-r/2})$ for $n \in \mathbb{N}$, then*

$$t_n = t_0 + \textstyle\sum_{m=1}^{r} n^{-m/2} Q_m^*(t_0) + o(n^{-r/2}) . \tag{9.10}$$

If (9.8) *holds with $o(n^{-r/2})$ replaced by $O(n^{-(r+1)/2})$, then* (9.9) *and* (9.10) *hold with $o(n^{-r/2})$ replaced by $O(n^{-(r+1)/2})$, unifomly for t in a compact subset.*

If the coefficients of $Q_1 \cdots, Q_r$, are considered as variables, these relations hold uniformly for coefficients belonging to compact sets.

LEMMA 8. *For $r \geqq 2$, $m = 1, \cdots, r$ let $R_m(t, \theta)$ be a polynomial in t with coefficients depending on θ, say $R_m(t, \theta) = \sum_{j=0}^{i(m)} r_{mj}(\theta) t^j$. Assume that r_{mj} is $(r - m)$-times differentiable and that the derivative of order $r - m$ fulfills a Lipschitz-condition. Assume that regularity conditions* (i)—(viii) *are fulfilled for $s = r + 1$.*

Then there exist polynomials in t, say $R_\mu^*(t, \theta)$, $\mu = 1, \cdots, r$ whose coefficients are rational functions of $r_{mj}^{(k)}(\theta)$ for $m = 1, \cdots, \mu$; $k = 0, 1, \cdots, \mu - m, j = 1, \cdots, i(m)$, such that uniformly in θ, τ on compact subsets of Θ and uniformly for $|t| \leqq \log n$,

$$P_\theta^{\mathbb{N}}\{\mathbf{x} \in X^{\mathbb{N}} : \tau + \textstyle\sum_{m=1}^{r} n^{-m/2} R_m(t, \tau) < \Theta_n(\mathbf{x})\} \Delta \{\mathbf{x} \in X^{\mathbb{N}} : \tau < \theta_n(x) + \sum_{m=1}^{r} n^{-m/2} R_m^*(t, \theta_n(\mathbf{x}))\} \leqq o(n^{-(r-1)/2}) . \tag{9.11}$$

We have in particular (with $R_m' \equiv (\partial/\partial\theta) R_m$ etc.)

$$R_1^* = -R_1$$

$$R_2^* = R_1 R_1' - R_2$$

$$R_3^* = -R_1 R_1'^2 - \tfrac{1}{2} R_1^2 R_1'' + R_1' R_2 - R_3 .$$

The following lemma implies that the set $\{(x_1, \cdots, x_n) : \sum_{i=1}^{n} h(x_i, \theta, \tau) \geqq \lambda_{n,\theta}\}$ renders an approximately most powerful critical region if $h(\cdot, \theta, \tau)$ is "close" to $1^{(1)}(\cdot, \theta)$. (For applications see Theorems 3(i), 3(ii) and Proposition 1.)

LEMMA 9. *Assume that the likelihood contrast function fulfills regularity conditions* (iv), (va), *that condition* (xi) *is fulfilled and that* $1^{(1)}$ *fulfills Condition* L_3. *Let* $h(\ ,\theta,\tau): x \to \mathbb{R}$, $\theta, \tau \in \Theta$, *be a family of* $\mathscr{A}$*-measurable functions fulfilling Condition* C_*. *Assume that for each* $\theta \in \Theta$ *there exists a neighborhood* U_θ *of* θ *and an* $\mathscr{A}$*-measurable function* $k(\cdot, \theta)$ *such that for all* $x \in X$, $\tau \in U_\theta$

$$|h(x, \theta, \tau) - 1^{(1)}(x, \theta)| \leqq |\tau - \theta| k(x, \theta)$$

and

$$\sup_{\tau \in U_\theta} E(k(\cdot, \tau)^4) < \infty .$$

Assume, finally, that for some sequence $\tau_n \in \Theta$ *with* $n^{\frac{1}{2}}|\tau_n - \theta| \to 0$, *some sequence* $\lambda_{n,\theta}$ *and some* $\alpha \in (0, 1)$ *we have uniformly on some compact* $K \subset \Theta$:

$$(9.12) \qquad P_\theta^{\mathrm{N}}\{\mathbf{x} \in X^{\mathrm{N}} : \textstyle\sum_{i=1}^n h(x_i, \theta, \tau_n) \geqq \lambda_{n,\theta}\} = \alpha + o(n^{-\frac{1}{2}}) .$$

Then uniformly for $\theta \in K$, $|t| \leqq \log n$

$$(9.13) \qquad P^{\mathrm{N}}_{\theta+n^{-\frac{1}{2}}t}\{\mathbf{x} \in X^{\mathrm{N}} : \textstyle\sum_{i=1}^n h(x_i, \theta, \tau_n) \geqq \lambda_{n,\theta}\} = H_{n,\alpha}(t, \theta) + o(n^{-\frac{1}{2}})$$

where $H_{n,\alpha}(t, \theta)$ *is defined by* (4.6).

If $(n^{\frac{1}{2}}|\tau_n - \theta|)_{n \in \mathrm{N}}$ *remains bounded, then the assertion holds with* $o(n^{-\frac{1}{2}})$ *replaced* $O(n^{-1})$ *uniformly for* θ *and* t *on compact subsets.*

PROOF. (i) Let

$$(9.14) \qquad \begin{aligned} \mu_\delta(\theta, \tau) &\equiv E_\delta(h(\cdot, \theta, \tau)) \\ \sigma_\delta(\theta, \tau) &\equiv (E_\delta((h(\cdot, \theta, \tau) - \mu_\delta(\theta, \tau))^2))^{\frac{1}{2}} \\ \lambda_\delta(\theta, \tau) &\equiv \sigma_\delta(\theta, \tau)^{-3} E_\delta((h(\cdot, \theta, \tau) - \mu_\delta(\theta, \tau))^3) . \end{aligned}$$

By Lemma 4 and Lemma 7(i) we have uniformly for $\theta \in K$, $u \in \mathbb{R}$:

$$\begin{aligned} P_\theta^{\mathrm{N}}\{\mathbf{x} \in X^{\mathrm{N}} : n^{-\frac{1}{2}}\sigma_\theta(\theta, \tau_n)^{-1} &\textstyle\sum_{i=1}^n (h(x_i, \theta, \tau_n) - \mu_\theta(\theta, \tau_n)) \\ &> u - \tfrac{1}{6} n^{-\frac{1}{2}} \lambda_\theta(\theta, \tau_n)(1 - u^2)\} \\ &= 1 - \Phi(u) + o(n^{-\frac{1}{2}}) . \end{aligned}$$

Together with (9.12) this implies by Lemma 7 (ii) uniformly for $\theta \in K$:

$$\lambda_{n,\theta} = n\mu_\theta(\theta, \tau_n) - \sigma_\theta(\theta, \tau_n)[n^{\frac{1}{2}} N_\alpha + \tfrac{1}{6}\lambda_\theta(\theta, \tau_n)(1 - N_\alpha^2)] + o(1) .$$

Hence,

$$(9.15) \qquad \begin{aligned} &P^{\mathrm{N}}_{\theta+n^{-\frac{1}{2}}t}\{\mathbf{x} \in X^{\mathrm{N}} : \textstyle\sum_{i=1}^n h(x_i, \theta, \tau_n) \geqq \lambda_{n,\theta}\} \\ &\quad = P^{\mathrm{N}}_{\theta+n^{-\frac{1}{2}}t}\{\mathbf{x} \in X^{\mathrm{N}} : n^{-\frac{1}{2}}\sigma_{\theta+n^{-\frac{1}{2}}t}(\theta, \tau_n)^{-1} \textstyle\sum_{i=1}^n (h(x_i, \theta, \tau_n) \\ &\qquad - \mu_{\theta+n^{-\frac{1}{2}}t}(\theta, \tau_n)) > \sigma_{\theta+n^{-\frac{1}{2}}t}(\theta, \tau_n)^{-1} \\ &\qquad \times [n^{\frac{1}{2}}(\mu_\theta(\theta, \tau_n) - \mu_{\theta+n^{-\frac{1}{2}}t}(\theta, \tau_n)) \\ &\qquad - \sigma_\theta(\theta, \tau_n)(N_\alpha + \tfrac{1}{6}n^{-\frac{1}{2}}\lambda_\theta(\theta, \tau_n)(1 - N_\alpha^2))] + o(n^{-\frac{1}{2}})\} . \end{aligned}$$

(ii) For notational convenience let $g(\cdot, \theta, \tau) \equiv h(\cdot, \theta, \tau) - 1^{(1)}(\cdot, \theta)$, $\theta, \tau \in \Theta$, and $G(\theta, \tau) \equiv E_\theta(g(\cdot, \theta, \tau))$.

Now we shall obtain asymptotic expansions for $\mu_{\theta+n^{-\frac{1}{2}}t}$, $\sigma_{\theta+n^{-\frac{1}{2}}t}$, and $\lambda_{\theta+n^{-\frac{1}{2}}t}$. Starting from

$$p(x, \theta + n^{-\frac{1}{2}}t) = p(x, \theta) + n^{-\frac{1}{2}}tp^{(1)}(x, \theta) + \tfrac{1}{2}n^{-1}t^2p^{(2)}(x, \theta + \delta n^{-\frac{1}{2}}t), \quad \delta \in (0, 1)$$

and using

$$\frac{p^{(2)}(x, \tau)}{p(x, \tau)} = 1^{(1)}(x, \tau)^2 + 1^{(2)}(x, \tau)$$

we obtain with $\delta \in (0, 1)$

$$\frac{p(x, \theta + n^{-\frac{1}{2}}t)}{p(x, \theta)} = 1 + n^{-\frac{1}{2}}t1^{(1)}(x, \theta) + \tfrac{1}{2}n^{-1}t^2(1^{(1)}(x, \theta + \delta n^{-\frac{1}{2}}t)^2 + 1^{(2)}(x, \theta + \delta n^{-\frac{1}{2}}t)) \frac{p(x, \theta + \delta n^{-\frac{1}{2}}t)}{p(x, \theta)}.$$

Hence

$$\Big|h(x, \theta, \tau_n) \frac{p(x, \theta + n^{-\frac{1}{2}}t)}{p(x, \theta)} - [1^{(1)}(x, \theta) + g(x, \theta, \tau_n) + n^{-\frac{1}{2}}t1^{(1)}(x, \theta)^2 + n^{-\frac{1}{2}}t1^{(1)}(x, \theta)g(x, \theta, \tau_n) + \tfrac{1}{2}n^{-1}t^2(1^{(1)}(x, \theta)^3 + 1^{(1)}(x, \theta)1^{(2)}(x, \theta))]\Big|$$
$$\leqq n^{-1}\frac{t^2}{2}\Big(|1^{(1)}(x, \theta)||1^{(1)}(x, \theta + \delta n^{-\frac{1}{2}}t)^2 - 1^{(1)}(x, \theta)^2|$$
$$+ |1^{(1)}(x, \theta)||1^{(2)}(x, \theta + \delta n^{-\frac{1}{2}}t) - 1^{(2)}(x, \theta)|$$
$$+ |1^{(1)}(x, \theta)||1^{(1)}(x, \theta + \delta n^{-\frac{1}{2}}t)|^2 \Big|\frac{p(x, \theta + \delta n^{-\frac{1}{2}}t)}{p(x, \theta)} - 1\Big|$$
$$+ |1^{(1)}(x, \theta)||1^{(2)}(x, \theta + \delta n^{-\frac{1}{2}}t)| \Big|\frac{p(x, \theta + \delta n^{-\frac{1}{2}}t)}{p(x, \theta)} - 1\Big|$$
$$+ |g(x, \theta, \tau_n)||1^{(1)}(x, \theta + \delta n^{-\frac{1}{2}}t)|^2 \frac{p(x, \theta + \delta n^{-\frac{1}{2}}t)}{p(x, \theta)}$$
$$+ |g(x, \theta, \tau_n)||1^{(2)}(x, \theta + \delta n^{-\frac{1}{2}}t)| \frac{p(x, \theta + \delta n^{-\frac{1}{2}}t)}{p(x, \theta)}$$
$$\leqq n^{-\frac{3}{2}}t^3R(x, \theta) + n^{-1}t^2|\tau_n - \theta|S(x, \theta)$$

where $\sup_{\theta \in K} E_\theta(R(\cdot, \theta)) < \infty$ and $\sup_{\theta \in K} E_\theta(S(\cdot, \theta)) < \infty$.

Hence we obtain uniformly for $\theta \in K$, $|t| \leqq \log n$

$$\mu_{\theta+n^{-\frac{1}{2}}t}(\theta, \tau_n) = G(\theta, \tau_n) + n^{-\frac{1}{2}}t(L_2(\theta) + E_\theta(1^{(1)}(\cdot, \theta)g(\cdot, \theta, \tau_n))) + \tfrac{1}{2}n^{-1}t^2(L_3(\theta) + L_{11}(\theta)) + o(n^{-1}).$$

By the same type of argument we obtain uniformly for $\theta \in K$, $|t| \leqq \log n$

$$\sigma_{\theta+n^{-\frac{1}{2}}t}(\theta, \tau_n) = L_2(\theta)^{\frac{1}{2}}[1 + L_2(\theta)^{-1}(E_\theta(1^{(1)}(\cdot, \theta)g(\cdot, \theta, \tau_n)) + \tfrac{1}{2}n^{-\frac{1}{2}}tL_3(\theta))] + o(n^{-\frac{1}{2}})$$
$$\lambda_{\theta+n^{-\frac{1}{2}}t}(\theta, \tau_n) = L_2(\theta)^{-\frac{3}{2}}L_3(\theta) + o(1).$$

(iii) Using the relations obtained in (ii), (9.15) can now be evaluated as follows:

For notational convenience let

$$w \equiv (N_\alpha + tL_2(\theta)^{\frac{1}{2}}) + \tfrac{1}{2}n^{-\frac{1}{2}}L_2(\theta)^{-\frac{1}{2}}(t^2L_{11}(\theta) - tN_\alpha L_2(\theta)^{-\frac{1}{2}}L_3(\theta) + \tfrac{1}{3}(1 - N_\alpha{}^2)L_2(\theta)^{-1}L_3(\theta)) .$$

Using the relations derived in (ii) we obtain that uniformly for $\theta \in K$, $|t| \leqq \log n$

$$\begin{aligned}(9.16)\qquad &\sigma_{\theta+n^{-\frac{1}{2}}t}(\theta, \tau_n)^{-1}[n^{\frac{1}{2}}(\mu_\theta(\theta, \tau_n) - \mu_{\theta+n^{-\frac{1}{2}}t}(\theta, \tau_n)) \\ &\quad - \sigma_\theta(\theta, \tau_n)(N_\alpha + \tfrac{1}{6}n^{-\frac{1}{2}}\lambda_\theta(\theta, \tau_n)(1 - N_\alpha{}^2))] \\ &= -w + o(n^{-\frac{1}{2}}) .\end{aligned}$$

By Lemma 4, (9.15) and (9.16) together imply

$$\begin{aligned}(9.17)\qquad &P^{\mathrm{N}}_{\theta+n^{-\frac{1}{2}}t}\{\mathbf{x} \in X^{\mathrm{N}} : \textstyle\sum_{i=1}^n h(x_i, \theta, \tau_n) \geqq \lambda_{n,\theta}\} \\ &= \Phi(w) - n^{-\frac{1}{2}}\varphi(w)\tfrac{1}{6}\lambda_{\theta+n^{-\frac{1}{2}}t}(\theta, \tau_n)(1 - w^2) + o(n^{-\frac{1}{2}}) .\end{aligned}$$

Since uniformly for $\theta \in K$, $|t| \leqq \log n$

$$\begin{aligned}\Phi(w) &= \Phi(N_\alpha + tL_2(\theta)^{\frac{1}{2}}) + \tfrac{1}{2}n^{-\frac{1}{2}}L_2(\theta)^{-\frac{1}{2}}(t^2L_{11}(\theta) - tN_\alpha L_2(\theta)^{-\frac{1}{2}}L_3(\theta) \\ &\quad + \tfrac{1}{3}(1 - N_\alpha{}^2)L_2(\theta)^{-1}L_3(\theta))\varphi(N_\alpha + tL_2(\theta)^{\frac{1}{2}}) + o(n^{-\frac{1}{2}}) \\ \varphi(w)(1 - w^2) &= \varphi(N_\alpha + tL_2(\theta)^{\frac{1}{2}})(1 - (N_\alpha + tL_2(\theta)^{\frac{1}{2}})^2) + o(1)\end{aligned}$$

we obtain from (4.6) uniformly for $\theta \in K$, $|t| \leqq \log n$:

$$\Phi(w) - n^{-\frac{1}{2}}\varphi(w)\tfrac{1}{6}\lambda_{\theta+n^{-\frac{1}{2}}t}(\theta, \tau_n)(1 - w^2) = H_{n,\alpha}(t, \theta) + o(n^{-\frac{1}{2}}) .$$

Together with (9.17) this proves the assertion (9.13).

10. Proofs.

Proof of Theorem 1. (i) $|t| > (4(s - 2) \log n)^{\frac{1}{2}}$. We shall give a proof for the case $t > (4(s - 2) \log n)^{\frac{1}{2}}$. The proof for $t < -(4(s - 2) \log n)^{\frac{1}{2}}$ runs similarly.

We have

$$\begin{aligned}(10.1)\qquad &\left|P_\theta{}^{\mathrm{N}}\left\{\mathbf{x} \in X^{\mathrm{N}} : n^{\frac{1}{2}}\frac{\theta_n(\mathbf{x}) - \theta}{\beta(\theta)} < t\right\} - \Phi(t) - \varphi(t)\textstyle\sum_{m=1}^{s-2} n^{-m/2}A_m(t, \theta)\right| \\ &\leqq P_\theta{}^{\mathrm{N}}\left\{\mathbf{x} \in X^{\mathrm{N}} : n^{\frac{1}{2}}\frac{\theta_n(\mathbf{x}) - \theta}{\beta(\theta)} \geqq t\right\} \\ &\quad + \Phi(-t) + \varphi(t)\textstyle\sum_{m=1}^{s-2} n^{-m/2}|A_m(t, \theta)| .\end{aligned}$$

By Lemma 3,

$$\begin{aligned}P_\theta{}^{\mathrm{N}}\left\{\mathbf{x} \in X^{\mathrm{N}} : n^{\frac{1}{2}}\frac{\theta_n(\mathbf{x}) - \theta}{\beta(\theta)} \geqq t\right\} &\leqq P_\theta{}^{\mathrm{N}}\left\{\mathbf{x} \in X^{\mathrm{N}} : n^{\frac{1}{2}}\frac{\theta_n(\mathbf{x}) - \theta}{\beta(\theta)}\right. \\ &\qquad\left.\geqq (4(s - 2) \log n)^{\frac{1}{2}}\right\} = o(n^{-(s-2)/2}) .\end{aligned}$$

Using Feller **1**, page 166, Lemma 2, we obtain

$$\Phi(-t) \leqq \Phi(-(4(s - 2) \log n) = o(n^{-(s-2)/2}) .$$

Furthermore,

$$t^p\varphi(t) \leqq (4(s-2)\log n)^{p/2}\varphi((4(s-2)\log n)^{\frac{1}{2}}) = o(n^{-(s-2)/2}) .$$

Hence the right side of (10.1) is $o(n^{-(s-2)/2})$.

(ii) $|t| \leqq (4(s-2)\log n)^{\frac{1}{2}}$. By Lemma 6 it suffices to prove

$$(10.2) \qquad P_\theta^N\{\mathbf{x}\in X^N: \textstyle\sum_{i=1}^n f^{(1)}(x_i, \theta + n^{-\frac{1}{2}}t\beta(\theta)) > 0\} = \Phi(t) + \varphi(t)\sum_{m=1}^{s-1} n^{-m/2}A_m(t,\theta) + o(n^{-(s-2)/2}) .$$

Let $\rho_k(\theta, v) \equiv E_\theta((f^{(1)}(\cdot, \theta + v\beta(\theta)))^k)$ and $\sigma(\theta, v) \equiv (\rho_2(\theta, v) - \rho_1(\theta, v)^2)^{\frac{1}{2}}$.

By Lemma 4 we have uniformly for $\theta\in K$, $|v| \leqq c_K$:

$$(10.3) \qquad \begin{aligned} P_\theta^N\{\mathbf{x}\in X^N: \textstyle\sum_{i=1}^n f^{(1)}(x_i, \theta + v\beta(\theta)) > 0\} &= P_\theta^N\{\mathbf{x}\in X^N: n^{-\frac{1}{2}} \textstyle\sum_{i=1}^n \sigma(\theta, v)^{-1}(-f^{(1)}(x_i, \theta + v\beta(\theta)) \\ &\qquad + \rho_1(\theta, v)) < n^{\frac{1}{2}}\rho_1(\theta, v)\sigma(\theta, v)^{-1}\} \\ &= \Phi(n^{\frac{1}{2}}\rho_1(\theta, v)\sigma(\theta, v)^{-1}) \\ &\qquad + \varphi(n^{\frac{1}{2}}\rho_1(\theta, v)\sigma(\theta, v)^{-1}) \textstyle\sum_{m=1}^{s-2} n^{-m/2} \\ &\qquad \times Q_m(n^{\frac{1}{2}}\rho_1(\theta, v)\sigma(\theta, v)^{-1}, \theta, \theta + v\beta(\theta)) + o(n^{-(s-2)/2}) . \end{aligned}$$

Now the right side of (10.3) considered as a function of v is expanded in a Taylor series about $v = 0$ with $(s-1)$ terms. Inserting $v = n^{-\frac{1}{2}}t$ and rearranging the terms according to their order in $n^{-\frac{1}{2}}$ leads to the assertion.

Starting from the Taylor expansion

$$f^{(1)}(x, \theta + v\beta(\theta))^k = \sum_{m=0}^{s-k} \frac{v^m}{m!}\beta(\theta)^m \frac{\partial^m}{\partial\theta^m}(f^{(1)}(x, \theta)^k) + o(v^{s-k}, x, \theta)$$

we obtain by condition (x) that uniformly on compact subsets of Θ,

$$(10.4) \qquad \rho_k(\theta, v) = \sum_{m=0}^{s-k} \frac{v^m}{m!}\beta(\theta)^m\rho_{km}(\theta) + o(v^{s-k}) .$$

From this we obtain uniformly on compact subsets of Θ,

$$\sigma(\theta, v)^{-1} = \sum_{m=0}^{s-2} v^m\sigma_m(\theta) + o(v^{s-2}) ,$$

with

$$(10.5) \qquad \begin{aligned} \sigma_0 &= \rho_{20}^{-\frac{1}{2}} \\ \sigma_1 &= -\tfrac{1}{2}\rho_{21}\rho_{20}^{-1}\rho_{11}^{-1} \\ \sigma_2 &= \tfrac{3}{8}\rho_{21}^2\rho_{20}^{-\frac{3}{2}}\rho_{11}^{-2} - \tfrac{1}{4}\rho_{22}\rho_{20}^{-\frac{1}{2}}\rho_{11}^{-2} + \tfrac{1}{2}\rho_{20}^{-\frac{1}{2}} . \end{aligned}$$

Furthermore,

$$(10.6) \qquad n^{\frac{1}{2}}\rho_1(\theta, n^{-\frac{1}{2}}t)\sigma(\theta, n^{-\frac{1}{2}}t)^{-1} = \sum_{m=0}^{s-2} n^{-m/2}t^{m+1}r_m(\theta) + n^{-(s-2)/2}t^{s-2}z_\theta(n^{-\frac{1}{2}}t) ,$$

where z_θ denotes a generic function with $\lim_{u\to 0}\sup_{\theta\in K} z_\theta(u) = 0$ and

$$(10.7) \qquad \begin{aligned} r_0 &= 1 \\ r_1 &= \tfrac{1}{2}\rho_{12}\rho_{20}^{\frac{1}{2}}\rho_{11}^{-2} - \tfrac{1}{2}\rho_{21}\rho_{20}^{-\frac{1}{2}}\rho_{11}^{-1} \\ r_2 &= \tfrac{1}{6}\rho_{13}\rho_{20}\rho_{11}^{-3} + \tfrac{3}{8}\rho_{21}^2\rho_{20}^{-1}\rho_{11}^{-2} - \tfrac{1}{4}\rho_{22}\rho_{11}^{-2} - \tfrac{1}{4}\rho_{12}\rho_{21}\rho_{11}^{-3} + \tfrac{1}{2} . \end{aligned}$$

From (10.6) we obtain for $\theta \in K$ and $n \geqq n_K$ (say):

$$(10.8) \qquad \Phi(n^{\frac{1}{2}}\rho_1(\theta, n^{-\frac{1}{2}}t)\sigma(\theta, n^{-\frac{1}{2}}t)^{-1}) = \Phi(t) + \varphi(t) \textstyle\sum_{m=1}^{s-2} n^{-m/2} R_m(t, \theta) + o(n^{-(s-2)/2}) z_\theta(n^{-\frac{1}{2}}t)$$

with

$$(10.9) \qquad R_1(t, \cdot) = r_1 t^2$$
$$R_2(t, \cdot) = r_2 t^3 - \tfrac{1}{2} r_1^2 t^5 .$$

By expanding φ in a Taylor series about t with $(s-2)$ terms and using (10.6) we obtain for $\theta \in K$ and $n \geqq n_K$ (say):

$$(10.10) \qquad \varphi(n^{\frac{1}{2}}\rho_1(\theta, n^{-\frac{1}{2}}t)\sigma(\theta, n^{-\frac{1}{2}}t)^{-1}) = \varphi(t) \textstyle\sum_{m=0}^{s-3} n^{-m/2} S_m(t, \theta) + n^{-(s-3)/2} p_\theta(t) \varphi(\tfrac{1}{2}t) z_\theta(n^{-\frac{1}{2}}t)$$

where $p_\theta(t)$ is a polynomial in t with coefficients depending on θ and being bounded on compact subsets of Θ. (This particular form of the remainder is needed since (10.10) will be multiplied with polynomials in t, say $q_\theta(t)$, and we shall need that $\sup_{t \in \mathbb{R}} \sup_{\theta \in K} q_\theta(t) p_\theta(t) \varphi(\frac{1}{2}t) < \infty$.)

Furthermore,

$$(10.11) \qquad S_0(t, \cdot) = 1 , \qquad S_1(t, \cdot) = -r_1 t^3 .$$

From Lemma 4, (10.6) and (10.10) we easily obtain for $\theta \in K$, $n \geqq n_K$:

$$(10.12) \qquad \varphi(n^{\frac{1}{2}}\rho_1(\theta, n^{-\frac{1}{2}}t)\sigma(\theta, n^{-\frac{1}{2}}t)^{-1}) \times Q_m(n^{\frac{1}{2}}\rho_1(\theta, n^{-\frac{1}{2}}t)\sigma(\theta, n^{-\frac{1}{2}}t)^{-1}, \theta + n^{-\frac{1}{2}}t\beta(\theta)) = \varphi(t) \textstyle\sum_{k=0}^{s-m-2} n^{-(k/2)} P_{mk}(t, \theta) + o(n^{-(s-m-2)/2}) z_\theta(n^{-\frac{1}{2}}t)$$

with

$$(10.13) \qquad \begin{aligned} P_{10}(t, \cdot) &= -\tfrac{1}{6}\rho_{30}\rho_{20}^{-\frac{3}{2}}(1 - t^2) \\ P_{11}(t, \cdot) &= (\tfrac{1}{2} + \tfrac{1}{4}\rho_{21}\rho_{30}\rho_{20}^{-2}\rho_{11}^{-1} - \tfrac{1}{6}\rho_{31}\rho_{20}^{-1}\rho_{11}^{-1})t \\ &\quad + (\tfrac{1}{2}\rho_{30}\rho_{20}^{-\frac{3}{2}} r_1 + \tfrac{1}{6}\rho_{31}\rho_{20}^{-1}\rho_{11}^{-1} - \tfrac{1}{4}\rho_{21}\rho_{30}\rho_{20}^{-2}\rho_{11}^{-1} - \tfrac{1}{2})t^3 \\ &\quad - \tfrac{1}{6}\rho_{30}\rho_{20}^{-\frac{3}{2}} r_1 t^5 \\ P_{20}(t, \cdot) &= (-\tfrac{3}{8} - \tfrac{5}{24}\rho_{20}^2\rho_{20}^{-3} + \tfrac{1}{8}\rho_{40}\rho_{20}^{-2})t \\ &\quad + (\tfrac{1}{8} - \tfrac{1}{24}\rho_{40}\rho_{20}^{-2} + \tfrac{5}{36}\rho_{30}^2\rho_{20}^{-3})t^3 - \tfrac{1}{72}\rho_{30}^2\rho_{20}^{-3} t^5 . \end{aligned}$$

(10.3), (10.8), and (10.12) together imply

$$(10.14) \qquad P_\theta^{\mathrm{N}}\{\mathbf{x} \in X^{\mathrm{N}} : \textstyle\sum_{i=1}^{n} f^{(1)}(x_i, \theta + n^{-\frac{1}{2}}t\beta(\theta)) > 0\} = \Phi(t) + \varphi(t) \sum_{m=1}^{s-2} n^{-m/2}(R_m(t, \theta) + \sum_{k=1}^{m} P_{k,m-k}(t, \theta)) + o(n^{-(s-2)/2}) .$$

Hence (3.1) holds with $A_m \equiv R_m + \sum_{k=1}^{m} P_{k,m-k}$.

The assertion of Theorem 1 now follows from (10.7), (10.9) and (10.13).

PROOF OF THEOREM 2. By Theorem 1 the following relation holds uniformly on compact subsets of Θ and uniformly for $t \in \mathbb{R}$:

$$P_\theta^{\mathrm{N}}\left\{\mathbf{x} \in X^{\mathrm{N}} : n^{\frac{1}{2}} \frac{\theta_n(\mathbf{x}) - \theta}{\beta(\theta)} < t\right\} = \Phi(t) + \varphi(t) \textstyle\sum_{m=1}^{s-2} n^{-m/2} A_m(t, \theta) + o(n^{-(s-2)/2}) .$$

By Lemma 7 we obtain that uniformly on compact subsets of Θ and uniformly for $|t| \leqq \log n$:

$$P_\theta^{\mathbb{N}}\left\{\mathbf{x} \in X^{\mathbb{N}} : n^{\frac{1}{2}} \frac{\theta_n(\mathbf{x}) - \theta}{\beta(\theta)} < t + \sum_{m=1}^{s-2} n^{-m/2} B_m(t, \theta)\right\} = \Phi(t) + o(n^{-(s-2)/2}) .$$

Applied for $t = -N_\alpha$ this yields the assertion of Theorem 2.

PROOF OF THEOREM 3. (i) Let $h(\cdot, \theta, \tau)$ be defined by

$$\begin{aligned} h(\cdot, \theta, \tau) &\equiv (\theta - \tau)^{-1}(1(\cdot, \theta) - 1(\cdot, \tau)) \qquad && \tau \neq \theta \\ &\equiv 1^{(1)}(\cdot, \tau) && \theta = \tau \end{aligned}$$

and let $\mu_\delta(\theta, \tau)$, $\sigma_\delta(\theta, \tau)$, $\lambda_\delta(\theta, \tau)$ be defined by (9.14).

By Lemma 4 and Lemma 7 we have uniformly for $\theta \in K$, $|t| \leqq \log n$, $|u| \leqq \log n$

$$\begin{aligned} P_\theta^{\mathbb{N}}\{\mathbf{x} \in X^{\mathbb{N}} : \sigma_\theta(\theta, \theta + n^{-\frac{1}{2}}t)^{-1} n^{-\frac{1}{2}} \textstyle\sum_{i=1}^n (h(x_i, \theta, \theta + n^{-\frac{1}{2}}t) \\ - \mu_\theta(\theta, \theta + n^{-\frac{1}{2}}t)) > u - \tfrac{1}{6} n^{-\frac{1}{2}} \lambda_\theta(\theta, \theta + n^{-\frac{1}{2}}t)(1 - u^2)\} \\ = 1 - \Phi(u) + o(n^{-\frac{1}{2}}) , \end{aligned}$$

i.e.

$$\begin{aligned} r_{K,n} \equiv \sup_{\theta \in K} \sup_{|u| \leqq \log n} \sup_{|t| \leqq \log n} |P_\theta^{\mathbb{N}}\{\mathbf{x} \in X^{\mathbb{N}} : \sigma_\theta(\theta, \theta + n^{-\frac{1}{2}}t)^{-1} n^{-\frac{1}{2}} \\ \times \textstyle\sum_{i=1}^n (h(x_i, \theta, \theta + n^{-\frac{1}{2}}t) - \mu_\theta(\theta, \theta + n^{-\frac{1}{2}}t)) > u \\ - \tfrac{1}{6} n^{-\frac{1}{2}} \lambda_\theta(\theta, \theta + n^{-\frac{1}{2}}t)(1 - u^2)\} - (1 - \Phi(u))| \end{aligned}$$

fulfills $\lim_{n\to\infty} n^{\frac{1}{2}} r_{K,n} = 0$.

Let $\alpha_{K,n} \equiv \sup_{\theta \in K} E_\theta^{\mathbb{N}}(\varphi_{\theta,n})$. By assumption,

$$\lim_{n\to\infty} n^{\frac{1}{2}}(\alpha_{K,n} - \alpha) = 0 .$$

Let

$$u_{K,n} \equiv N_{1-\alpha_{K,n}-r_{K,n}} .$$

Since

$$\begin{aligned} n^{\frac{1}{2}}|u_{K,n} - N_{1-\alpha}| &= n^{\frac{1}{2}}|N_{1-\alpha_{K,n}-r_{K,n}} - N_{1-\alpha}| \\ &\leqq n^{\frac{1}{2}}|\alpha_{K,n} - \alpha + r_{K,n}| \qquad \text{for all} \quad n \geqq n_K \quad \text{(say), we have} \\ u_{K,n} &= -N_\alpha + o(n^{-\frac{1}{2}}) , \end{aligned}$$

so that $\{u_{K,n} : n \in \mathbb{N}\}$ is bounded. Hence we have for all $\theta \in K$, $|t| \leqq \log n$

$$\begin{aligned} P_\theta^{\mathbb{N}}\{\mathbf{x} \in X^{\mathbb{N}} : \sigma_\theta(\theta, \theta + n^{-\frac{1}{2}}t)^{-1} n^{-\frac{1}{2}} \textstyle\sum_{i=1}^n (h(x_i, \theta, \theta + n^{-\frac{1}{2}}t) \\ - \mu_\theta(\theta, \theta + n^{-\frac{1}{2}}t)) > u_{K,n} - n^{-\frac{1}{2}}\tfrac{1}{6}\lambda_\theta(\theta, \theta + n^{-\frac{1}{2}}t)(1 - u_{K,n}^2)\} \\ \geqq 1 - \Phi(u_{K,n}) - r_{K,n} = \alpha_{K,n} \geqq E_\theta^{\mathbb{N}}(\varphi_{\theta,n}) . \end{aligned}$$

In other words, with

$$\lambda_{n,\alpha} \equiv n\mu_\theta(\theta, \theta + n^{-\frac{1}{2}}t) + \sigma_\theta(\theta, \theta + n^{-\frac{1}{2}}t)(n^{\frac{1}{2}} u_{K,n} - \tfrac{1}{6}\lambda_\theta(\theta, \theta + n^{-\frac{1}{2}}t)(1 - u_{K,n}^2))$$

we have for $\theta \in K$, $0 < t < \log n$

$$P_\theta^{\mathbb{N}}\{\mathbf{x} \in X^{\mathbb{N}} : \textstyle\sum_{i=1}^n h(x_i, \theta, \theta + n^{-\frac{1}{2}}t) \geqq \lambda_{n,\theta}\} \geqq E_\theta^{\mathbb{N}}(\varphi_{\theta,n}) .$$

Since for $t > 0$ any critical region $\{\mathbf{x} \in X^{\mathbb{N}} : \sum_{i=1}^n h(x_i, \theta, \theta + n^{-\frac{1}{2}}t) > \lambda\}$ is most powerful for testing the hypothesis θ against the alternative $\theta + n^{-\frac{1}{2}}t$, this implies for $\theta \in K$, $0 < t \leqq \log n$:

(10.15) $$P^{\mathbf{N}}_{\theta+n^{-\frac{1}{2}}t}\{\mathbf{x} \in X^{\mathbf{N}}: \textstyle\sum_{i=1}^n h(x_i, \theta, \theta + n^{-\frac{1}{2}}t) \geqq \lambda_{n,\theta}\} \geqq E^{\mathbf{N}}_{\theta+n^{-\frac{1}{2}}t}(\varphi_{\theta,n}) .$$

Since $1^{(1)}$ fulfills Condition L_4, the function h fulfills the conditions assumed in Lemma 9. Since

$$P_\theta{}^{\mathbf{N}}\{\mathbf{x} \in X^{\mathbf{N}}: \textstyle\sum_{i=1}^n h(x_i, \theta, \theta + n^{-\frac{1}{2}}t) \geqq \lambda_{n,\theta}\} = \alpha + o(n^{-\frac{1}{2}}) ,$$

Lemma 9 may be applied for $\tau_n = \theta + n^{-\frac{1}{2}}t$ to obtain

$$P^{\mathbf{N}}_{\theta+n^{-\frac{1}{2}}t}\{\mathbf{x} \in X^{\mathbf{N}}: \textstyle\sum_{i=1}^n h(x_i, \theta, \theta + n^{-\frac{1}{2}}t) \geqq \lambda_{n,\theta}\} = H_{n,\alpha}(t, \theta) + o(n^{-\frac{1}{2}}) .$$

Together with (10.15) this proves (4.5) for $0 < t < \log n$.

Since $t \geqq \log n$ implies $\inf_{\theta \in K} (N_\alpha + tL_2(\theta)^{\frac{1}{2}}) > (2 \log n)^{\frac{1}{2}}$ for all $n \geqq n_K$ (say), we have by Lemma 2 in Feller **1**, page 166, for all $\theta \in K$, $t \geqq \log n$

$$\Phi(N_\alpha + tL_2(\theta)^{\frac{1}{2}}) > \Phi((2 \log n)^{\frac{1}{2}}) \geqq 1 - n^{-1} ,$$

so that (4.5) holds trivially in this case.

The proof for $t < 0$ runs similarly.

(ii) We have

$$P_\theta{}^{\mathbf{N}}\{\mathbf{x} \in X^{\mathbf{N}}: \theta_n(\mathbf{x}) \geqq G^{(s)}_{n,\alpha}(\theta)\} = \alpha + o(n^{-\frac{1}{2}}) .$$

As $n^{\frac{1}{2}}(G^{(s)}_{n,\alpha}(\theta) - \theta)$ is bounded, we obtain by Lemma 6:

$$P_\theta{}^{\mathbf{N}}\{\mathbf{x} \in X^{\mathbf{N}}: \textstyle\sum_{i=1}^n 1^{(1)}(x_i, G^{(s)}_{n,\alpha}(\theta)) \geqq 0\} = \alpha + o(n^{-\frac{1}{2}}) .$$

Since $1^{(1)}$ fulfills Condition L_4, the conditions of Lemma 9 are fulfilled for the function $h(\cdot, \theta, \tau) = 1^{(1)}(\cdot, \tau)$ and $\lambda_{n,\theta} = 0$. Hence we obtain uniformly for $\theta \in K$, $|t| \leqq \log n$:

$$P^{\mathbf{N}}_{\theta+n^{-\frac{1}{2}}t}\{\mathbf{x} \in X^{\mathbf{N}}: \textstyle\sum_{i=1}^n 1^{(1)}(x_i, G^{(s)}_{n,\alpha}(\theta)) \geqq 0\} = H_{n,\alpha}(t, \theta) + o(n^{-\frac{1}{2}}) .$$

Applying Lemma 6 once again we obtain uniformly for $\theta \in K$, $|t| \leqq \log n$:

$$P^{\mathbf{N}}_{\theta+n^{-\frac{1}{2}}t}\{\mathbf{x} \in X^{\mathbf{N}}: \theta_n(\mathbf{x}) \geqq G^{(s)}_{n,\alpha}(\theta)\} = H_{n,\alpha}(t, \theta) + o(n^{-\frac{1}{2}}) .$$

We now consider the case $\log n \leqq t \leqq c_K n^{\frac{1}{2}}$. As $n^{\frac{1}{2}}(G^{(s)}_{n,\alpha}(\theta) - \theta)$ and $\beta(\theta + n^{-\frac{1}{2}}t)$ are bounded (uniformly for $\theta \in K$, $\log n \leqq t \leqq c_K n^{\frac{1}{2}}$), we obtain for $t \geqq \log n$, $\theta \in K$, $n \geqq n_K$,

$$\beta(\theta + n^{-\frac{1}{2}}t)^{-1}(n^{\frac{1}{2}}(G^{(s)}_{n,\alpha}(\theta) - \theta) - t) \leqq -2(\log n)^{\frac{1}{2}} .$$

This implies

$$\begin{aligned} P^{\mathbf{N}}_{\theta+n^{-\frac{1}{2}}t}&\{\mathbf{x} \in X^{\mathbf{N}}: \theta_n(\mathbf{x}) \geqq G^{(s)}_{n,\alpha}(\theta)\} \\ &\geqq P^{\mathbf{N}}_{\theta+n^{-\frac{1}{2}}t}\{\mathbf{x} \in X^{\mathbf{N}}: n^{\frac{1}{2}}\beta(\theta + n^{-\frac{1}{2}}t)^{-1} \\ &\qquad \times (\theta_n(\mathbf{x}) - (\theta + n^{-\frac{1}{2}}t)) \geqq -2(\log n)^{\frac{1}{2}}\} \\ &= 1 - o(n^{-\frac{1}{2}}) \end{aligned}$$

with the last inequality following from Lemma 3.

Since $H_{n,\alpha}(t, \theta) = 1 + o(n^{-\frac{1}{2}})$ for $\theta \in K$ and $\log n \leqq t$ by Feller **1**, page 166, Lemma 2, this implies the assertion for $\log n \leqq t \leqq c_K n^{\frac{1}{2}}$.

The proof for $-c_K n^{\frac{1}{2}} \leqq t \leqq -\log n$ runs similarly.

PROOF OF PROPOSITION 1. By Lemma 4 and Lemma 7 we have uniformly for $\theta \in K$:

$$P_\theta^{\mathbb{N}}\{\mathbf{x} \in X^{\mathbb{N}} : \sum_{i=1}^n 1^{(1)}(x_i, \theta) \geqq -n^{\frac{1}{2}} N_\alpha L_2(\theta)^{\frac{1}{2}} + L_2(\theta)^{-1} L_3(\theta) \tfrac{1}{6}(N_\alpha{}^2 - 1)\} = \alpha + o(n^{-\frac{1}{2}}) .$$

Lemma 9 (applied for $h(\cdot, \theta, \tau) = 1^{(1)}(\cdot, \theta)$) implies the assertion.

PROOF OF THEOREM 4. Follows from Theorem 2 and Lemma 8.

PROOF OF THEOREM 5. (i) Let $\varphi_{n,\theta} \equiv 1_{\{\mathbf{x} \in X^{\mathbb{N}} : \theta \notin C_n(\mathbf{x})\}}$. We have uniformly on compact subsets of Θ:

$$E_\theta^{\mathbb{N}}(\varphi_{n,\theta}) = \alpha + o(n^{-\frac{1}{2}}) .$$

For compact $K \subset \Theta$ choose a compact $K_0 \subset \Theta$ such that $K \subset K_0{}^0$. Since $\theta \in K$, $0 < t < \log n$ implies $\theta - n^{-\frac{1}{2}}t \in K_0$ for all $n \geqq n_K$ (say), we have uniformly for $\theta \in K$, $0 < t < \log n$

$$E^{\mathbb{N}}_{\theta - n^{-\frac{1}{2}}t}(\varphi_{n, \theta - n^{-\frac{1}{2}}t}) = \alpha + o(n^{-\frac{1}{2}}) .$$

Hence we obtain from Theorem 3 (i) (applied for $\theta - n^{-\frac{1}{2}}t$ instead of θ)

$$E_\theta^{\mathbb{N}}(\varphi_{n, \theta - n^{-\frac{1}{2}}t}) \leqq H_{n,\alpha}(t, \theta - n^{-\frac{1}{2}}t) + o(n^{-\frac{1}{2}}) .$$

From this (5.4) follows easily for $0 < t < \log n$. The remaining cases may be treated in an analogous way as in the proof of Theorem 3 (i).

(ii) By Theorem 3 (ii) we have uniformly for $\theta \in K$, $|t| < c_K n^{\frac{1}{2}}$

$$P_\theta^{\mathbb{N}}\{\mathbf{x} \in X^{\mathbb{N}} : \theta_n(\mathbf{x}) > G^{(s)}_{n,\alpha}(\theta - n^{-\frac{1}{2}}t)\} = H_{n,\alpha}(t, \theta - n^{-\frac{1}{2}}t) + o(n^{-\frac{1}{2}}) .$$

By Lemma 8,

$$P_\theta^{\mathbb{N}}\{\mathbf{x} \in X^{\mathbb{N}} : \theta_n(\mathbf{x}) > G^{(s)}_{n,\alpha}(\theta - n^{-\frac{1}{2}}t)\} = P_\theta^{\mathbb{N}}\{\mathbf{x} \in X^{\mathbb{N}} : F^{(s)}_{n,\alpha}(\theta_n(\mathbf{x})) > \theta - n^{-\frac{1}{2}}t\} + o(n^{-(s-2)/2}) .$$

These two relations together imply the assertion.

PROOF OF THEOREM 6. (i) Let $C_n(\mathbf{x}) \equiv \{\theta \in \Theta : \theta \geqq T_n(\mathbf{x})\}$. According to the assumption on T_n, $n \in \mathbb{N}$, we have by (6.4)

$$P_\theta^{\mathbb{N}}\{\mathbf{x} \in X^{\mathbb{N}} : \theta \in C_n(\mathbf{x})\} \geqq \tfrac{1}{2} - o(n^{-\frac{1}{2}}) .$$

Hence by Theorem 5 (i),

$$P_\theta^{\mathbb{N}}\{\mathbf{x} \in X^{\mathbb{N}} : \theta - n^{-\frac{1}{2}}t' \in C_n(\mathbf{x})\} \geqq 1 - H_{n,\frac{1}{2}}(t', \theta - n^{-\frac{1}{2}}t') - o(n^{-\frac{1}{2}}) .$$

A straightforward computation shows that uniformly for $\theta \in K$, $0 < t' < \log n$,

$$1 - H_{n,\frac{1}{2}}(t', \theta - n^{-\frac{1}{2}}t') = \Phi(-t' L_2(\theta)^{\frac{1}{2}}) + n^{-\frac{1}{2}} \varphi(t' L_2(\theta)^{\frac{1}{2}}) t'^2 L_2(\theta)^{-\frac{1}{2}} (\tfrac{1}{3} L_3(\theta) + \tfrac{1}{2} L_{11}(\theta)) + o(n^{-\frac{1}{2}}) .$$

Hence we obtain uniformly for $\theta \in K$, $0 < t' < \log n$,

$$P_\theta{}^N\{\mathbf{x} \in X^N : T_n(\mathbf{x}) \leqq \theta - n^{-\frac{1}{2}}t'\}$$
$$\geqq \Phi(-t'L_2(\theta)^{\frac{1}{2}}) + n^{-\frac{1}{2}}\varphi(t'L_2(\theta)^{\frac{1}{2}})$$
$$\times\, t'^2 L_2(\theta)^{-\frac{1}{2}}(\tfrac{1}{3}L_3(\theta) + \tfrac{1}{2}L_{11}(\theta)) - o(n^{-\frac{1}{2}})\,.$$

The case $t' \geqq \log n$ may be treated in the usual way.

Since Property (6.3) is dual to (6.4) we obtain the dual result

$$P_\theta{}^N\{\mathbf{x} \in X^N : T_n(\mathbf{x}) \geqq \theta + n^{-\frac{1}{2}}t'')$$
$$\geqq \Phi(-t''L_2(\theta)^{\frac{1}{2}}) - n^{-\frac{1}{2}}\varphi(t''L_2(\theta)^{\frac{1}{2}})$$
$$\times\, t''^2 L_2(\theta)^{-\frac{1}{2}}(\tfrac{1}{3}L_3(\theta) + \tfrac{1}{2}L_{11}(\theta)) - o(n^{-\frac{1}{2}})\,.$$

The proof of (ii) and (iii) is obvious.

Acknowledgment. The author wishes to thank Mr. P. J. Bickel whose suggestions on the general orientation of the paper as well as on several details led to substantial improvements. He furthermore wishes to thank Mr. R. Michel who carefully read several versions of the manuscript and suggested various modifications, and the *Deutsche Forschungsgemeinschaft* which enabled the cooperation with Mr. Michel by a grant. He finally wishes to thank Mr. D. A. Sprott for helpful discussions, and Mr. Th. Pfaff, Mr. J. Weber, and Mr. H. Wegner for computations in connection with the examples in Section 7.

REFERENCES

Bartlett, M. S. (1953). Approximate confidence intervals. *Biometrika* **40** 12–19.

Bol'shev, L. N. (1959). On transformations of random variables. *Theor. Probability. Appl.* **4** 129–141.

Borges, R. (1971). Derivation of normalizing transformations with an error of order $1/n$. *Sankhyā* **33** 441–459.

Chibisov, D. M. (1972a). On the normal approximation for a certain class of statistics. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* **1** 153–174.

Chibisov, D. M. (1972b). Asymptotic expansions for maximum likelihood estimates. *Theor. Probability Appl.* **17** 387–388 (in Russian).

Chung, K. L. (1951). The strong law of large numbers. *Proc. Second Berkeley Symp. Math. Statist. Prob.* 341–352.

Cramér, H. (1970). *Random Variables and Probability Distributions.* Cambridge Univ. Press.

Esséen, C. G. (1945). Fourier analysis of distribution functions. *Acta Math.* **77** 1–125.

Feller, W. (1957). *An Introduction to Probability Theory and its Applications*, **1**. Wiley, New York.

Feller, W. (1966). *An Introduction to Probability Theory and its Applications*, **2**. Wiley, New York.

Gnedenko, S. V. and Kolmogorov, A. N. (1954). *Limit Distributions for Sums of Independent Random Variables.* Addison-Wesley, Cambridge.

Haldane, B. S. and Smith, S. M. (1956). The sampling distribution of a maximum-likelihood estimate. *Biometrika* **43** 96–103.

Hodges, J. L., Jr. and Lehmann, E. L. (1970). Deficiency. *Ann. Math. Statist.* **41** 783–801.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under non-standard conditions. *Proc. Fifth Berkeley Symp. Math. Statist Prob.* **1** 221–233.

Huzurbazar, V. S. (1947). The likelihood equation, consistency and the maxima of the likelihood function. *Ann. Eugenics* **14** 185–200.

KENDALL, M. G. and STUART, A. S. (1961). *The Advanced Theory of Statistics*, **2**. Griffin, London.

LEHMANN, E. L. (1959). *Testing Statistical Hypotheses*. Wiley, New York.

MICHEL, R. (1972). Results on probabilities of moderate deviations. To appear in *Ann. Probability*.

MICHEL, R. (1973). Bounds for the efficiency of approximately median unbiased estimates. To appear in *Metrika*.

MICHEL, R. and PFANZAGL, J. (1971). The accuracy of the normal approximation for minimum contrast estimates. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **18** 73–84.

MITROFANOVA, N. M. (1967). An asymptotic expansion for the maximum likelihood estimate of a vector parameter. *Theor. Probability Appl.* **12** 364–372.

NEYMAN, J. (1959). Optimal asymptotic test of composite statistical hypotheses. *Probability and Statistics*, (U. Grenander, ed.). Almquist, Stockholm. 213–234.

PFANZAGL, J. (1970). Median unbiased estimates for m.l.r.—families. *Metrika* **15** 30–39.

PFANZAGL, J. (1971). The Berry-Esséen bound for minimum contrast estimates. *Metrika* **17** 82–91.

PFANZAGL, J. (1972). On median unbiased estimates. *Metrika* **18** 154–173.

PFANZAGL, J. (1973). The accuracy of the normal approximation for estimates of vector parameters. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **25** 171–198.

RAO, C. R. (1962). Efficient estimates and optimum inference procedures in large samples. *J. Roy. Statist. Soc.* **24** 46–63.

RAO, C. R. and POTI, S. J. (1946). On locally most powerful tests when alternatives are one-sided. *Sankhyā* **7** 439.

RAO, R. R. (1961). On the central limit theorem in R_k. *Bull. Amer. Math. Soc.* **67** 359–361.

SHENTON, L. R. and BOWMAŇ, K. (1963). Higher moments of a maximum-likelihood estimate. *J. Roy. Statist. Soc. Ser. B* **25** 305–317.

WALD, A. (1941 a). Asymptotically most powerful tests of statistical hypotheses. *Ann. Math. Statist.* **12** 1–19.

WALD, A. (1941 b). Some examples of asymptotically most powerful tests. *Ann. Math. Statist.* **12** 396–408.

WASOW, W. (1956). On the asymptotic transformation of certain distributions into the normal distribution. *Proc. Symp. Appl. Math.* **6** 252–259.

WELCH, B. L. (1939). On the distribution of maximum likelihood. *Biometrika* **31** 187–190.

WELCH, B. L. (1965). On comparisons between confidence point procedures in the case of a single parameter. *J. Roy. Statist. Soc. B* **27** 1–8.

WILKS, S. S. (1963). *Mathematical Statistics*. Wiley, New York.

MATHEMATISCHES INSTITUT
DER UNIVERSITAT ZU KÖLN
5 KÖLN
WEYERTAL 86
WEST GERMANY