CHAPTER 2

Edgeworth Expansions, Curved Exponentials and Fisher Consistent Estimates

2.1. Edgeworth expansions for functions of sample mean. To explain what the expansions for asymptotic mean and variance mean, we need the concept of Edgeworth expansions.

Let X_i, $i = 1, 2, \ldots, n$, be m-dimensional, i.i.d., $f_1(\cdot), \ldots, f_k(\cdot)$, k real valued (measurable) functions on R^m and

$$Z_j = \big(f_1(X_j), f_2(X_j), \ldots, f_k(X_j)\big)$$

with finite mean $\mu = E(Z_j)$ and finite positive definite dispersion matrix $[\sigma_{jj'}]$. The ith coordinates of Z_j and μ are Z_{ij} and μ_i. Let

$$\overline{Z} = \frac{1}{n}\sum_1^n Z_j, \qquad T \equiv T_n = H(\overline{Z}),$$

where H is a real valued function which is $(s-1)$ times continuously differentiable in a neighborhood of μ, $s \geq 2$.

By Taylor expansion around μ,

$$\begin{aligned} H(\overline{Z}) &= H(\mu) + \sum l_i(\overline{Z}_i - \mu_i) + \sum l_{ii'}(\overline{Z}_i - \mu_i)(\overline{Z}_{i'} - \mu_{i'}) \\ &\quad + \cdots + \sum l_{i_1, i_2, \ldots, i_{s-1}}(\overline{Z}_{i_1} - \mu_{i_1})(\overline{Z}_{i_2} - \mu_{i_2})\cdots \\ &\qquad\qquad \times (\overline{Z}_{i_{s-1}} - \mu_{i_{s-1}}) + R_n \qquad \text{(2.1a)} \\ &\underset{\text{def}}{=} H_1(\overline{Z}) + R_n, \end{aligned}$$

where

$$l_i = \left.\frac{\partial H(Z)}{\partial z_i}\right|_\mu, \qquad l_{ii'} = \left.\frac{\partial^2 H}{\partial z_i\, \partial z_{i'}}\right|_\mu, \quad \text{and so on,}$$

and $R_n = o(\|\bar{Z} - \mu\|^{s-1})$. The order of the remainder R_n can be strengthened to $o(\|\bar{Z} - \mu\|^{s})$ if we assume H is s times continuously differentiable. Clearly, this implies

$$(2.1b)\quad \begin{aligned} R_n &= o_p(n^{-(s-1)/2}) \quad \text{if } H \text{ is } (s-1) \text{ times continuously differentiable} \\ &= O_p(n^{-s/2}) \quad \text{if } H \text{ is } s \text{ times continuously differentiable.} \end{aligned}$$

Let

$$W_n = \sqrt{n}\left(H(\bar{Z}) - H(\mu)\right),$$
$$W_n' = \sqrt{n}\left(H_1(\bar{Z}) - H(\mu)\right).$$

Applying the central limit theorem to the first term in the expansion for W_n or W_n', and noting that the subsequent terms are $o_p(1)$, it is easy to see W_n and W_n' are A.N. $(0, b_2)$, where $b_2 = \sum l_i l_{i'} \sigma_{ii'}$. We will always assume the l_i's are not all zero so that, $[\sigma_{ii'}]$ being positive definite, $b_2 > 0$.

Our object is to have an expansion of $P\{W_n \in B\}$, which is valid, that is, correct, up to $o(n^{-(s-2)/2})$, for all Borel sets B.

If $s = 2$, we have asymptotic normality of W_n, which leads to first order calculations. If $s = 3$, we have a one-term Edgeworth expansion, corresponding to second order calculations and if $s = 4$, we have a two-term Edgeworth expansion, corresponding to third order calculations. In statistical applications of higher order asymptotics we never need to go beyond $s = 4$.

We need another technical result to motivate how the expansion is calculated. It turns out that under the assumption of $(s - 1)$ times continuous differentiability of H, Condition D, which will be discussed a little later, and appropriate moment assumptions,

$$P\{W_n \in B\} - P\{W_n' \in B\} = o(n^{-(s-2)/2}).$$

The following simple lemma makes it plausible.

LEMMA 2.1. *Let* $Y_n = Y_n' + R_n'$ *and assume*

$$(2.2a)\quad P\left\{R_n' = o(n^{-(s-2)/2})\right\} = 1 - o(n^{-(s-2)/2})$$

and, uniformly in y,

$$(2.2b)\quad \begin{aligned} P\{Y_n' < y\} = A_1(y) &+ A_2(y)/\sqrt{n} + \cdots + A_{s-2}(y)/(\sqrt{n})^{s-2} \\ &+ o(n^{-(s-2)/2}), \end{aligned}$$

where each $A_i(\cdot)$ *is continuously differentiable. Then*

$$(2.3)\quad \begin{aligned} P\{Y_n < y\} = A_1(y) &+ A_2(y)/\sqrt{n} + \cdots + A_{s-2}(y)/(\sqrt{n})^{s-2} \\ &+ o(n^{-(s-2)/2}). \end{aligned}$$

By way of proof, note that for any $\varepsilon > 0$,

$$P\left\{|R_n' > \varepsilon^{-(s-2)/2}\right\} = o(n^{-(s-2)/2}).$$

Hence

$$P\{Y'_n \le y - \epsilon n^{-(s-2)/2}\} + o(n^{-(s-2)/2}) \le P\{Y_n \le y\} \le P\{Y'_n \le y + \epsilon n^{-(s-2)/2}\} + o(n^{-(s-2)/2}).$$

Now use (2.2b) for $y \pm \epsilon n^{-(s-2)/2}$ and the fact that ϵ may be made arbitrarily small.

Note (2.2a) is stronger than $R'_n = o_p(n^{-(s-2)/2})$ and it is the condition we need to have equality of the left-hand sides of (2.2b) and (2.3). Condition (2.2a) will often be used in the sequel tacitly. Let us verify it in the case of

$$W_n = W'_n + R'_n,$$

where $R'_n = \sqrt{n} R_n = o_p(n^{-(s-2)/2})$, from the Taylor expansion [see (2.1a) and (2.1b)]. Assume H is s times continuously differentiable and Z_j has finite sth order moments. Under the second assumption,

$$P\{\|\bar{Z} + \mu\| > \log n/\sqrt{n}\} = o(n^{-(s-2)/2}) \tag{2.4}$$

by a result of von Bahr; see Bhattacharya and Ghosh (1978). Under s times continuous differentiability of H, we have, with a suitable $c > 0$,

$$|R'_n| = |\sqrt{n} R_n| < c(\log n)^{s/2}/(n)^{(s-1)/2} \tag{2.5}$$

on the set $\|\bar{Z} - \mu\| \le \log n/\sqrt{n}$. The condition (2.2a) now follows from (2.4) and (2.5).

We now get back to Edgeworth expansion of $P\{W'_n \in B\}$, hoping by the above argument that it will also be Edgeworth expansion of $P\{W_n \in B\}$.

ASSUMPTION B. H is $(s-1)$ times continuously differentiable in a neighborhood of μ, and Z_j has finite sth order (absolute) moments, $[\sigma_{ii'}]$ is positive definite and $l_i = (\partial H(Z_i)/\partial Z_{ij}|_{\mu}$, $i = 1, \dots, k$, are not all zero.

Since W'_n is a polynomial in $(\bar{Z} - \mu)$, its moments and hence its cumulants are relatively easy to calculate. Let these moments and cumulants calculated up to $o(n^{-(s-2)/2})$ be denoted by $m_{r,n}$, $\kappa_{r,n}$. Our Assumption B suffices for these calculations. These moments and cumulants are often also called the moments and cumulants of W_n, calculated by the delta method, that is, essentially by Taylor expansion. This terminology is somewhat misleading for the following reason. The moments of W'_n are not, in general, good approximations to the moments of W_n, which are very sensitive to the tail of W_n; rather, it will turn out they are the moments of a distribution which approximates the distribution of W_n up to $o(n^{-(s-2)/2})$.

The κ's have expansions of the following kind. Let $[\cdot]$ denote the integral part and let ( ) denote a suitable coefficient free of n:

$$\begin{aligned}
\kappa_{1,n} &= (\ )n^{-1/2} + (\ )n^{-3/2} + \cdots + (\ )n^{-[(s-3)/2]+1/2},\\
\kappa_{2,n} &= b_2 + (\ )n^{-1} + \cdots + (\ )n^{-[(s-2)/2]},\\
\kappa_{r,n} &= (\ )n^{-(r-2)/2} + \cdots + (\ )n^{[(s-3)/2]+1/2} \quad \text{if } r \text{ is odd}\\
&= (\ )n^{-(r-2)/2} + \cdots + (\ )n^{-[(s-2)/2]} \quad \text{if } r \text{ is even.}
\end{aligned} \tag{2.6}$$

For $s = 4$, $\kappa_{1,n}$, $\kappa_{3,n}$ and $\kappa_{4,n}$ each has one term and $\kappa_{2,n}$ has two terms. Also $\kappa_{5,n}$ onward are $o(n^{-1})$. Note that, in general, $\kappa_{r,n} = o(n^{-(s-2)/2})$ for $r > s$ and hence is negligible. This is a nontrivial fact; see Bhattacharya and Ghosh (1978) or Bhattacharya and Denker (1990). Finally, as noted before, $b_2 = \sum l_i l_{i'} \sigma_{ii'}$ is the variance of the limiting normal distribution of W_n' or W_n.

For fixed real t, one has formally

$$\log E(e^{itW_n'}) = \sum_{r=1}^{s} \kappa_{r,n}(it)^r/r! + o(n^{-(s-2)/2}),$$

so that

$$E(e^{itW_n'}) = e^{-t^2/2}[1 + (\ )n^{-1/2} + (\ )n^{-1} + \cdots + (\ )n^{-(s-2)/2}] + o(n^{-(s-2)/2}), \tag{2.7}$$

where the coefficient of $n^{-r/2}$ is a polynomial $\pi_r(it)$. It is easy to prove by repeated integration by parts that

$$\pi_r(it)e^{-t^2/2} = \int_{-\infty}^{\infty} e^{ity}\left(\pi_r(-D)e^{-y^2/2b_2}\right) dy, \tag{2.8}$$

where $D = d/dy$ and $\pi_r(-D)f(y)$ is the result obtained by operating the polynomial in $(-D)$ on f. So inverting term by term on the right-hand side of (2.7) and using (2.8), we get

$$\begin{aligned} E(e^{itW_n'}) &= \text{c.f. of} \sum_{r=0}^{s-2} \{\pi_r(-D)\phi(y|0, b_2)\}n^{-r/2} + o(n^{-(s-2)/2}) \\ &\underset{\text{def}}{=} \text{c.f. of } \psi_{s,n} + o(n^{-(s-2)/2}), \end{aligned}$$

where c.f. denotes characteristic function and $\phi(y|0, b_2)$ is the normal density with mean 0 and variance b_2. The expression

$$\begin{aligned} \psi_{s,n} &= \sum_{0}^{s-2} \pi_r(-D)\phi(y|0, b_2)n^{-r/2} \\ &= \phi(y|0, b_2)\sum(\ )n^{-r/2}, \end{aligned} \tag{2.9}$$

where the coefficient of $n^{-r/2}$ is a polynomial in y, is called a formal, as distinct from a rigorous or valid, Edgeworth expansion of the density of W_n. We will call it a valid Edgeworth expansion if

$$P(W_n \in B) = \int_B \psi_{s,n}(y)\, dy + o(n^{-(s-2)/2}). \tag{2.10}$$

Often one requires (2.10) to hold uniformly in B.

To prove (2.10) one needs the following condition.

CONDITION D. For some positive integer M, the M-fold convolution of Z_1 has an absolutely continuous nonzero component.

An easily verifiable sufficient condition is provided in the following lemma. Recall that $Z_{i1} = f_i(X_1)$, where Z_{i1} is the ith component of the vector Z_1.

LEMMA 2.2. *Suppose X_1 has a probability density function (with respect to the Lebesgue measure) which is positive on some open ball B. Suppose also $f_1, \dots, f_k$ are continuously differentiable on B and $1, f_1, f_2, \dots, f_k$ are linearly independent (as elements of the linear space of continuous functions on B). Then Condition D holds with $M = k$.*

A proof is available in Bhattacharya and Ghosh (1978). It is essentially similar to a proof of a famous result of Dynkin on sufficiency.

THEOREM 2.1. *Suppose X_j's are i.i.d., $f_1, f_2, \dots, f_k$, Z_j, W_n and W_n' are defined as before and Assumption B and Condition D hold. Then*

$$\sup_B \left| P\{W_n \in B\} - \int_B \psi_{s,n}(y)\,dy \right| = o(n^{-(s-2)/2}), \tag{2.10a}$$

$$\sup_B \left| P\{W_n' \in B\} - \int_B \psi_{s,n}(y)\,dy \right| = o(n^{-(s-2)/2}), \tag{2.10b}$$

where the supremum is over all Borel sets.

Thus under the conditions of Theorem 2.1 we do have validity of the formal Edgeworth expansion for W_n and W_n'.

We first give two examples, and then, in the next two sections, remark on the proof and assumptions.

EXAMPLE 2.1 (Student's t). Assume X_j's are real valued and have positive density in some interval B. Consider Student's t, where

$$t = \bar{X}/s \quad \text{where } \bar{X} = \frac{1}{n}\sum_1^n X_j \text{ and } s^2 = \frac{1}{n}\sum X_i^2 - \bar{X}^2,$$

which may be written as $H(\bar{Z})$, where

$$Z_j = \big(f_1(X_j), f_2(X_j)\big) \equiv \big(X_j, X_j^2\big).$$

Clearly, the conditions of Lemma 2.1 hold. If X_1 has finite mean ν_1 and finite variance ν_2, then

$$\mu = \big(\nu_1, \nu_2 + \nu_1^2\big).$$

If further X_1 has finite fourth order moment, then $[\sigma_{ii'}]$ is finite and positive definite. H is infinitely differentiable in a neighborhood of μ. If X_1 has finite moment of order q, q even, then Theorem 2.1 applies with $s = q/2$. For $s = 3$, $\nu_1 = 0$, the expansion is given by

$$P\{\sqrt{n}t \le y\} = \Phi(y|0,1) + n^{-1/2}P_1(y)\phi(y|0,1),$$

where Φ and ϕ are the standard normal distribution and density functions, and

$$P_1(y) = \frac{1}{6}\frac{\nu_3}{\nu_2^{3/2}}\{2y^2 + 1\}.$$

Note that we need only third order moments to write down the one-term Edgeworth expansion but need six moments to prove validity by Theorem 2.1.

EXAMPLE 2.2. $X_j = (U_j, V_j)$, U_j, V_j real valued,

$$r = \text{sample correlation coefficient between } U \text{ and } V$$

$$= H(\bar{Z}) = \left(\overline{UV} - \bar{U}\bar{V}\right)/\left\{\overline{U^2} - (\bar{U})^2\right)\left(\overline{V^2} - (\bar{V})^2\right)\right\}^{1/2},$$

where $Z_j = (U_j, V_j, U_j^2, V_j^2, U_jV_j)$. Assume X_j has a p.d.f. which is positive on a ball B and has finite moments of order q, q even. Then Theorem 2.1 holds with $s = q/2$, since H may be differentiated as often as one wants in any neighborhood of $E(Z_j)$ that excludes zero as a denominator. Note that if X_j has a positive density in a ball B, then $E(U_j^2) - (E(U_j))^2 > 0$ and $E(V_j^2) - (E(V_j))^2 > 0$.

2.2. Remarks on the proof. The proof with s times continuous differentiability appears in Bhattacharya and Ghosh (1978). By paying more careful attention to the remainder in the Taylor expansion, Bhattacharya (1985) shows essentially the same proof works with $(s - 1)$ times continuous differentiability, that is, Assumption B. We provide a sketch of the main steps.

One begins by noting that the formal Edgeworth expansion for the vector $\sqrt{n}(\bar{Z} - \mu)$ is a valid one, by the results in Bhattacharya and Rao (1976). The proof of this fundamental fact is quite technical and involves Fourier arguments, a Berry–Esseen type argument and truncation to avoid making unnecessary moment assumptions. When $\bar{Z}$ is one dimensional, a proof is available in Feller (1966).

One then makes a change of variables,

$$\sqrt{n}(\bar{Z} - \mu) \to \left(\sqrt{n}\left(\bar{Z}_1 - \mu_1\right), \dots, \sqrt{n}\left(\bar{Z}_{p-1} - \mu_{p-1}\right), W_n\right) \equiv T,$$

and notes that this is a perturbation of a linear transformation since W_n is a perturbation of a linear function of $\sqrt{n}(\bar{Z} - \mu)$. This is used to show that the joint density of T may be taken as the product of a multivariate normal and a polynomial in T and $n^{-1/2}$. Essentially it shows T has an Edgeworth expansion. If $\bar{Z}_1, \dots, \bar{Z}_{p-1}$ are integrated out, one gets the marginal of W_n, which is also an Edgeworth expansion but the coefficient polynomials cannot be written down explicitly.

It remains to verify that this expansion agrees with the formal Edgeworth expansion. This is done by an indirect argument which shows the two expansions have the same moments of all order and hence must be identical.

2.3. Remarks on assumptions, extensions, background and other applications. For a general H, the moment assumptions just suffice for writing down the Edgeworth expansion. However, for special H, say, Student's t, the one-term Edgeworth expansion requires only finite third order moment of X_1, whereas Theorem 2.1 would require finiteness of twice as many moments. Hall (1987) shows third order moments suffice for the validity of the one-term Edgeworth expansion for Student's t. For relaxation of moment assumptions in a more general setting, see Bhattacharya and Ghosh (1988), and Babu and Bai (1992).

Condition D holds in all common examples where X_j is a continuous random variable or a random vector. On the other hand, Condition D as well as the conclusion of Theorem 2.1 fails when X is lattice valued, the case of greatest interest when X is discrete. To see this, take X_j to be binomial $P\{X_j = x\} = p^x q^{1-x}$, $x = 0, 1$, $p = 1/2$, say, $Z_j = X_j$, $H(\bar{Z}) = H(\bar{X}) = \bar{X}$. Then, using Stirling's approximation, it is easy to show $P(\bar{Z} = 1/2) \sim n^{-1/2}$, that is, the ratio of the two quantities has a positive limit. However if the $(s-2)$-term Edgeworth expansion holds for $s \geq 3$, $P(\bar{Z} = 1/2) = o(n^{-(s-2)/2}) = o(n^{-1/2})$, since the expansion puts zero mass on singletons.

Using results of Götze and Hipp (1978), one can show, without Condition D,

$$\int f \, dP = \int f \psi_{s,n}(y) \, dy + o(n^{-(s-2)/2}) \tag{2.10c}$$

for f belonging to a certain class of smooth functions with conditions on growth at infinity. See Bhattacharya and Ghosh (1978) and Ghosh, Sinha and Subramanyam (1979) for more details. Later we apply (2.10c) for expansion of asymptotic mean, variance and risk $E(l(\sqrt{n}(T_n - \theta))$ of an estimate $T_n = H(\bar{Z})$, in discrete cases.

Consider Borel sets B, satisfying

$$\int_{(\partial B)^\varepsilon} \Phi(dy|0,1) = O(\varepsilon) \quad \text{as } \varepsilon \downarrow 0, \tag{2.10d}$$

where ∂B is the boundary of B, $(\partial B)^\varepsilon$ is the set of all points within a distance of ε from ∂B and Φ is the standard normal. For any R^p, the measurable convex sets satisfy this condition. If we want Theorem 2.1 only for Borel sets satisfying (2.10d), we can relax Condition D to Condition C, Cramér's condition:

$$\limsup_{\|t\| \to \infty} \left| E\{\exp i\langle t, Z_j\rangle\} \right| < 1,$$

where $\langle t, Z_j\rangle$ is the scalar product. However, we do not know of any interesting statistical example where Condition D fails but Condition C holds. For a proof of this version of Theorem 2.1, see Bhattacharya and Ghosh (1978), Bhattacharya (1985) and Bhattacharya and Denker (1990).

Theorem 2.1 along with all the above remarks continue to hold if W_n is a random vector of dimension less than or equal to that of $\bar{Z}$; see Bhattacharya and Ghosh (1978).

If the conditional distribution of one component of Z_1, given the remaining components, satisfies Condition C, then, under slightly stronger moment assumptions, Bai and Rao (1991) show Theorem 2.1 holds for Borel sets satisfying (2.10d). Such examples occur naturally in the context of life testing and survival analysis, owing to the discreteness of one or more components due to censoring. For an early example of Edgeworth expansion in this setting, see Basu and Ghosh (1980). Presumably a p-dimensional analogue of Bai and Rao (1991) is easy to prove.

We end this section with a few remarks on the very extensive literature on Edgeworth expansions and applications to areas other than higher order efficiency.

The best treatment of Edgeworth expansion for the sample mean of independent random vectors is Bhattacharya and Rao (1976). Bhattacharya and Denker (1990) contains a very readable self-contained account under Condition C. The one-dimensional case, which is technically much simpler, can be found in Feller (1966). Wallace (1958) and Bickel (1974) contain excellent reviews with a focus on statistical applications. Both contain Theorem 2.1 or some version in the form of conjecture.

The most notable application of Edgeworth expansions, other than in the study of higher order efficiency of estimates or the [Hodges–Lehmann (1970)] deficiency of tests, is the proof of superiority of the bootstrap over the classical delta method. The pioneering paper in this context is Singh (1981), who shows that for Student's t the bootstrap does as well as the one-term Edgeworth expansion, if X_1 satisfies Condition C. Refinements or extensions are available in Babu and Singh (1984), Bhattacharya and Qumsiyeh (1989) and Hall (1986). An altogether different point of view on the superiority of the bootstrap is taken in Ghosh (1992).

For asymptotic expansions of sums of dependent r.v.'s the basic paper is Götze and Hipp (1983). For sums of the form $\Sigma f(X_i)$ or $\Sigma f(X_i, X_{i-1})$ for Markov chains, see Jensen (1986), which also contains a very readable introduction to the whole subject and a good overview.

2.4. Curved exponential families and Fisher consistent estimates. Begin with a k-parameter exponential density or probability function

$$p(x|\beta) = d(\beta)\exp\left\{\sum_1^k \beta_i f_i(x)\right\}A(x), \tag{2.11}$$

where $\beta = (\beta_1, \dots, \beta_k)$ is an element of some open k-dimensional rectangle $V \subset R^k$, $f_1, f_2, \dots, f_k$ are real valued and $1, f_1, \dots, f_k$ are linearly independent in the sense

$$a_0 + \sum_1^k a_j f_j = 0 \quad \text{a.e. [under } p(x|\beta)]$$

implies all a's are zero.

Let Θ be an open interval in R and let

$$\theta \to (\beta_1(\theta), \beta_2(\theta), \dots, \beta_k(\theta)) \equiv \beta(\theta)$$

be a curve in V. Let

$$d(\beta(\theta)) \underset{\text{def}}{=} c(\theta).$$

Then

$$p(x|\theta) = c(\theta)\exp\left\{\sum \beta^i(\theta) f_i(x)\right\} A(x), \qquad \theta \in \Theta, \tag{2.12}$$

is a curved exponential. It reduces to a one parameter linear exponential if $\beta_i(\theta)$'s are all linear in θ.

Of course if Θ is a p-dimensional open rectangle contained in R^p, $p < m$, we have a p-dimensional curved exponential.

Such densities have often been used. The terminology is due to Efron (1975). The multinomial in Chapter 1 is a special case, with

$$\beta_i = \log\left\{\pi_i \Big/ \left(1 - \sum_1^k \pi_i\right)\right\}, \qquad f_i = I\{x = x_i\}. \tag{2.13}$$

EXAMPLE 2.3 (Berkson's logistic bioassay model). Suppose the probability of death at a given dose d_i (of a particular substance) is

$$\pi_i(\alpha, \beta) = (1 + e^{-(\alpha + \beta d_i)})^{-1}. \tag{2.14}$$

Suppose each of the k doses $d_1, \dots, d_k$ is given to n animals, different for different doses, and the number of survivors at each level noted. The joint p.f. is a product of k binomials, which may be regarded as an exponential family with

$$X_{ij} = 1 \quad \text{if the } j\text{th animal getting the } i\text{th dose } d_i \text{ dies}$$
$$= 0 \quad \text{otherwise.}$$

$f_i(X_j) = X_{ij}$, $\beta_i = \log \pi_i(1 - \pi_i)^{-1}$, $\theta = (\alpha, \beta)$. It is actually a linear exponential family.

EXAMPLE 2.4 (Behrens–Fisher). Let $X_j = (U_j, V_j)$ be i.i.d., where U_j and V_j are independent $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, respectively. Behrens and Fisher considered the problem of testing H_0: $\mu_1 = \mu_2$. Historically this was the first example which made clear the fundamental difference between Fisher's fiducial probability and Neyman's confidence coefficient. Let us assume H_0 is true and denote the common value of μ_i's as μ. Then we have a curved exponential with $\theta = (\mu, \sigma_1^2, \sigma_2^2)$, $f_1(X_j) = U_j$, $f_2(X_j) = U_j^2$, $f_3(X_j) = V_j$, $f_4(X_j) = V_j^2$. The dimension of the (minimal) sufficient statistic $(\Sigma_{j=1}^n f_i(X_j)$, $i = 1, 2, 3, 4)$ is 4 while the dimension of the parameter space is 3. The (minimal) sufficient statistic is not even boundedly complete, and Unni, in a thesis submitted to the Indian Statistical Institute in 1978, showed no function of θ, including its components, has a minimum variance unbiased estimate. However, TOE estimates for the components of θ, asymptotically unbiased up to $o(n^{-1})$, are easy to construct.

In this example, the curved exponential is obtained by imposing a polynomial relation on the natural parameters, where $\beta_1 = \mu_1/2\sigma_1^2$, $\beta_2 = -1/2\sigma_1^2$ and β_3, β_4 are similarly defined. Such curved exponentials, sometimes called algebraic exponential families, have a rich theory drawing on algebra and algebraic geometry. There is a beautiful theorem due to Kagan and Palamodov which completely characterizes all parametric functions for which a minimum variance unbiased estimate exists. A source for all this is Linnik (1967).

We return to the general curved exponential and let

$$\mu(\beta) = E(f \mid \beta). \tag{2.15}$$

Then for $\beta \in V$, μ is (real) analytic. So if we assume $\beta(\theta)$ is q times continuously differentiable, then

$$\mu(\beta(\theta)) \equiv \mu(\theta)$$

is also q times continuously differentiable. The same fact holds for $c(\theta)$.

Let

$$Z_j = \left(f_1(X_j), f_2(X_j), \dots, f_k(X_j)\right) \tag{2.16}$$

so that $\mu_i(\theta) = E(f_i|\theta) = E(Z_{ij}|\theta)$. To make the analogy with a multinomial clear, we will sometimes have the identification

$$\mu_i(\theta) = \pi_i(\theta), \qquad \overline{Z} = p, \qquad \overline{Z}_i = p_i. \tag{2.17}$$

DEFINITION 2.1. An estimate T_n is Fisher consistent (FC) if

$$T_n = H(\overline{Z}),$$

where H is a real valued function such that

$$H(\mu(\theta)) = \theta \qquad \forall \theta.$$

If $p(x|\beta)$ is probability density (rather than a probability function) and $A(x)$ is positive on an open interval, then Condition D holds with f_i's and Z_j's as defined in Section 2.1. All moments of Z_1 are finite. So if H is $(s-1)$ times continuously differentiable in a neighborhood of $\mu(\theta)$, the estimate $T_n = H(\overline{Z})$ has a valid Edgeworth expansion correct up to $o(n^{-(s-2)/2})$.

2.5. Maximum likelihood estimator for curved exponentials. We turn to the maximum likelihood estimate (mle) $\hat{\theta}_n$. For us it will mean a suitably chosen consistent solution of the likelihood equation based on n i.i.d. observations $X_1, X_2, \dots, X_n$.

Here the common density of p.f. is given by (2.12). We assume the β_i's are thrice continuously differentiable in θ. The likelihood equation is

$$\frac{\dot{C}(\theta)}{C(\theta)} + \sum_1^k \dot{\beta}(\theta)\overline{Z}_i = 0, \tag{2.18}$$

which has a solution $\theta = \theta_0$ if $\bar{Z}$ is replaced by $\mu(\theta_0)$. If we denote the left-hand side of (2.18) by $A(\theta, \bar{Z})$, then

$$\left.\frac{\partial A}{\partial \theta}\right|_{\theta_0, \mu(\theta_0)} = -(\text{Fisher information at } \theta_0) < 0. \tag{2.19}$$

It follows by the implicit function theorem that there exists a neighborhood N of $\mu(\theta_0)$ and $\delta > 0$ such that for $\bar{Z}$ in N, a thrice continuously differentiable solution of (2.18),

$$T_n = H(\bar{Z}), \qquad |T_n - \theta_0| < \delta,$$

can be found. This solution is unique. If θ is sufficiently close to θ_0, $|\theta - \theta_0| < \delta_1 < \delta$, then $\mu(\theta) \in N$, so that by uniqueness of H on N,

$$H(\mu(\theta)) = \theta.$$

It is this solution that we will call an mle and denote by $\hat{\theta}$ whenever we have a curved exponential. Of course this is a local construction. A global construction is provided in the next paragraph under a compactness assumption. The reader who is not interested in this may skip the next paragraph.

Suppose Θ has a compact closure, $\bar{\Theta}$, $\beta(\cdot)$ has a continuous extension on $\bar{\Theta}$, $\{\beta(\theta)\colon \theta \in \bar{\Theta}\} \subset V$, and we have the identifiability condition $\beta(\theta) \neq \beta(\theta')$, $\theta, \theta' \in \bar{\Theta}$, $\theta \neq \theta'$. Under these assumptions, we have enough uniformity to choose δ, δ_1 to be the same for all $\theta_0 \in \bar{\Theta}$, N a ball $B(\mu(\theta_0), \delta_2)$, δ_2 free of θ_0, $\delta_3 > 0$ such that $B(\mu(\theta), \delta_3) \subset B(\mu(\theta'), \delta_2)$ if $|\theta - \theta'| < \delta_1$. Moreover,

$$\delta_4 = \inf\{\|\mu(\theta) - \mu(\theta')\|;\ \theta, \theta' \in \bar{\Theta}, |\theta - \theta'| \geq \delta_1\} > 0.$$

Choose δ_5 to be smaller than δ_3 as well as $\delta_4/2$. Then two balls $B(\theta_i, \delta_5)$ and $B(\theta_j, \delta_5)$ intersect only if $|\theta_i - \theta_j| < \delta_1$ and, moreover, the union of both balls is contained in $N = B(\mu(\theta_i), \delta_2)$, so that H is unambiguously defined on their union. Cover the compact set $\{\mu(\theta);\ \theta \in \bar{\Theta}\}$ by the balls $B(\mu(\theta), \delta_5)$, find a finite subcover $B(\mu(\theta_i), \delta_5)$, $i = 1, 2, \ldots, m$, by the Heine–Borel theorem and define H unambiguously on $\bigcup_{i=1}^{m} B(\mu(\theta_i), \delta_5)$. The probability that $\bar{Z}$ falls outside this set is exponentially small, uniformly in $\theta \in \bar{\Theta}$, by an inequality of Chernoff (1952). Hence there we may assign arbitrary values in Θ to our mle. This completes the construction.

Since

$$\hat{\theta} = H(\bar{Z})$$

for $\bar{Z}$ in a neighborhood of $\mu(\theta_0)$ and H is thrice continuously differentiable, Theorem 2.1 applies with $s = 4$ whenever we have a continuous curved exponential. Of course if we assume $\beta(\theta)$ is s-times differentiable, the theorem would hold with that s.

2.6. Maximum likelihood estimator in the general regular case. Suppose Assumption A of Chapter 1 is strengthened so that $p(x|\theta)$ is five times continuously differentiable and holds with the third derivative replaced by the fifth derivative. Assume $E_{\theta_0}|M(X)|^4 < \infty$.

We will assume Condition D for Z_j, where $Z_j = (Z_{1i}, Z_{2i}, Z_{3i}, Z_{4i})$,

$$Z_{ij} = \left.\frac{d^i \log p(x_j|\theta)}{d\theta^i}\right|_{\theta_0}.$$

We also assume Z_j has finite fourth moments. In the following, $\bar{Z} = (1/n)\Sigma Z_j$.

We essentially reproduce the proof of Theorem 2.3 of Bhattacharya and Ghosh (1978), for $s = 4$.

Use a standard argument involving the sign change of a continuous function, or a fixed point theorem in the multiparameter case [see Bhattacharya and Ghosh (1978)] to prove that the likelihood equation has a solution which converges in probability to θ_0. Applying von Bahr's inequality on $\bar{Z}$, it is possible to ensure that, with P_{θ_0}-probability $1 - o(n^{-1})$, $\hat{\theta}$ satisfies the likelihood equation and lies in $(\theta_0 \pm \log n/\sqrt{n})$. It is this solution that we take as our mle. Of course there is a problem of identifying such a solution if the likelihood equation has multiple roots, since the true θ_0 will be unknown to the statistician. This problem can be resolved if we have a consistent estimate T_n such that T_n lies in $(\theta_0 \pm \log n/\sqrt{n})$ with P_{θ_0}-probability $1 - o(n^{-1})$. In this case, we may take the solution nearest to T_n. By the preceding reasoning, this solution, which is identifiable from the sample, will lie in $(\theta_0 \pm \log n/\sqrt{n})$ with P_{θ_0}-probability $1 - o(n^{-1})$.

Clearly, with $\hat{\theta}$ as above, with probability $1 - o(n^{-1})$,

$$0 = \bar{Z}_1 + \bar{Z}_2(\hat{\theta} - \theta_0) + \frac{\bar{Z}_3}{2!}(\hat{\theta} - \theta_0)^2 + \frac{\bar{Z}_4}{3}(\hat{\theta} - \theta_0)^3 + R_n,$$

where, as defined earlier,

$$\bar{Z}_i = \frac{1}{n}\frac{d^i \log p(x_1, x_2, \dots, x_n|\theta)}{d\theta^i}$$

and

$$R_n = \frac{1}{n}\left.\frac{d^5 \log p}{d\theta^5}\right|_{\theta'}(\hat{\theta} - \theta)^4 = O\left(\frac{\log n}{\sqrt{n}}\right)^4.$$

We rewrite this equation as

$$0 = A(\bar{Z}, \hat{\theta}) + R_n.$$

Note

$$0 = A(\mu(\theta_0), \theta_0)$$

and

$$\left.\frac{\partial A}{\partial \theta}\right|_{\theta_0, \mu(\theta_0)} = -(\text{Fisher information}) \neq 0.$$

Hence, as in the previous section,

$$0 = A(\bar{Z}, \tilde{\theta})$$

has a solution $\tilde{\theta} = H(\bar{Z})$, $|\tilde{\theta} - \theta| < \delta$, in a neighborhood of $\mu(\theta_0)$ and H is continuously differentiable to all order. Moreover, by definition of $\hat{\theta}$ and uniqueness of H,

$$\hat{\theta} = H\left(\bar{Z}_1 + R_n, \bar{Z}_2, \bar{Z}_3, \bar{Z}_4\right)$$

with P_{θ_0}-probability $1 - o(n^{-1})$. This implies, with probability $1 - o(n^{-1})$,

$$|\tilde{\theta} - \hat{\theta}| \leq K(\log n/\sqrt{n})^4$$

Also Theorem 2.1 applies to $\tilde{\theta}$. We now apply Lemma 2.1 to deduce that $P_{\theta_0}\{\sqrt{n}(\hat{\theta} - \theta_0) < y\}$ is given by the integral of the valid two-term Edgeworth expansion for $\sqrt{n}(\tilde{\theta} - \theta)$ up to $o(n^{-1})$.

The same method applies to M-estimates.

Note that we have not been able to prove whether $P_{\theta_0}\{\sqrt{n}(\hat{\theta} - \theta_0) \in B\}$, for arbitrary Borel sets, has an expansion.

We write down the two-term Edgeworth expansion for $\sqrt{n}(\hat{\theta} - \theta_0)$ explicitly below. Let the cumulants of $\sqrt{n}(\tilde{\theta} - \theta)$ found by the delta method be written as follows [up to $O(n^{-1})$]:

$$\begin{aligned} \text{first cumulant} &= k'_{11} n^{-1/2}, \\ \text{second cumulant} &= b_2 + k'_{22} n^{-1}, \\ \text{third cumulant} &= k'_{31} n^{-1/2}, \\ \text{fourth cumulant} &= k'_{41} n^{-1}. \end{aligned} \tag{2.20a}$$

Let

$$\begin{aligned} k_{11} &= k'_{11}/\sqrt{b_2}, \\ k_{22} &= k'_{22}/b_2, \\ k_{31} &= k'_{31}/b_2^{3/2}, \\ k_{41} &= k'_{41}/b_2^2. \end{aligned} \tag{2.20b}$$

Then

$$\begin{aligned} &P_{\theta_0}\left\{\sqrt{n}\left(\hat{\theta} - \theta_0\right) \leq y/\sqrt{I}(\theta_0)\right\} \\ &\quad = \Phi(y) + \Phi_1(y)/\sqrt{n} + \Phi_2(y)/n + o(n^{-1}), \end{aligned} \tag{2.21}$$

where Φ is the standard normal distribution function,

$$\Phi_1(y) = \int_{-\infty}^{y} \{k_{11} H_1(u) + k_{31} H_3(u)\} \phi(u)\, du, \tag{2.22a}$$

$$\begin{aligned} \Phi_2(y) = \int_{-\infty}^{y} \Big\{ & k_{22} H_2(u)/2 + k_{41} H_4(u)/24 \\ & + \tfrac{1}{2}\left(k_{11}^2 H_2(u) + k_{31}^2 H_6(u)\right)/36 \\ & + \tfrac{1}{3}\left(k_{11} k_{31} H_4(u)\right)\Big\} \phi(u)\, du \end{aligned} \tag{2.22b}$$

with $\phi(y)$ equal to the standard normal density and H_p, given by

$$H_p(y)\phi(y) = \left(-\frac{d}{dy}\right)^p \phi(y),$$

is the pth Hermite polynomial. Note that the above integrals can be evaluated by making use of

$$\int_{-\infty}^{y} H_p(u)\phi(u)\,du = -H_{p-1}(y)\phi(y).$$

2.7. Remarks on expansion of asymptotic mean, variance and risk. These expansions were introduced and used in Chapter 1 to define TOE estimates. They will appear frequently in the later chapters also. These expansions are exactly what one would get if one calculates $E(\sqrt{n}(T_n - \theta))$, $E(n(T_n - \theta)^2 - \{E(\sqrt{n}(T_n - \theta))\}^2)$ by the delta method, correct up to $o(n^{-1})$. This presupposes $\sqrt{n}(T_n - \theta)$ can be written as $\sqrt{n}(H(\bar{Z}) - \mu) + R'_N$ in a neighborhood of $\theta = H(\mu)$, R'_N satisfies the condition of Lemma 2.1 with $s = 4$, H is thrice continuously differentiable and $\bar{Z}$ has finite fourth order absolute moment. In Section 2.3 we have exhibited such a $\bar{Z}$ and H for the mle. In Section 2.2 this was done for Fisher consistent estimates.

As explained in Section 2.1, these expansions may not always provide asymptotic expansions of the moments they are supposed to approximate. What meaning can be assigned to them then? There are three slightly different answers.

Note that the expansions up to $o(n^{-(s-2)/2})$ of the moments of $\sqrt{n}(T_n - \theta)$ calculated by the delta method are exactly the moments of $\psi_{s,n}$ up to $o(n^{-(s-2)/2}$. This is easily verified by reversing the steps by which the formal Edgeworth expansion was derived. At least two of the interpretations are based on this fact.

The first and, to us, the most satisfying, interpretation is based on the assumption that $\psi_{s,n}$ is, in fact, a valid Edgeworth expansion for $s = 4$. The expansions of asymptotic bias and variance are then the exact bias and variance [up to $o(n^{-1})$] of $\psi_{4,n}$ which approximates the distribution of $\sqrt{n}(T_n - \theta)$ up to $o(n^{-1})$.

Unfortunately for discrete examples, like binomials or multinomials, even the one-term formal Edgeworth expansion is not valid. In such cases, we can regard the expansions as expansions for suitably truncated mean and variance of $\sqrt{n}(T_n - \theta)$ and apply [see (2.10c)] the results of Götze and Hipp (1978) on the validity of Edgeworth expansions for expectations of smooth functions. This is done in Ghosh, Sinha and Subramanyam (1979).

The third option, available for curved exponentials, is to use a somewhat different truncation used in Ghosh and Subramanyam (1974). Then it is almost trivial to check directly the validity of the expansions for the truncated mean and variance.

Similar remarks hold for $E(l(\sqrt{n}(T_n - \theta)))$, which we calculate as

$$\int l(y)\psi_{s,n}(y)\,dy + o(n^{-(s-2)/2}).$$

If l is only known to be bounded and measurable, only the first interpretation is available. If l is also smooth, the other two interpretations are also available. In particular, if l is the indicator of the complement of a symmetric interval, only the first interpretation is available; see Ghosh, Sinha and Wieand (1980).

In the study of higher order efficiency, we will often encounter estimates of the form $T_n' = T_n + c(T_n)/n$. If $\sqrt{n}(T_n - \theta)$ has a valid Edgeworth expansion up to $o(n^{-1})$ and c is twice differentiable at θ, then T_n' also has a valid Edgeworth expansion up to $o(n^{-1})$. This may be seen as follows. Note

$$\sqrt{n}(T_n' - \theta) = \sqrt{n}(T_n - \theta)\left(1 + \frac{c'(\theta)}{n}\right) + R_n,$$

where $|R_n|$ can be shown to satisfy the condition of Lemma 2.1 by making use of the Edgeworth expansion of $\sqrt{n}(T_n - \theta)$ up to $o(n^{-1})$, and the first term on the right-hand side, which is a linear function of $\sqrt{n}(T_n - \theta)$, is easily seen to have a valid Edgeworth expansion. It follows by Lemma 2.1 that $\sqrt{n}(T_n' - \theta)$ has a valid Edgeworth expansion also. It turns out that by imitating the proof of Theorem 2.1, we can show this under the weaker assumption that c is once continuously differentiable. It is easy to show the valid Edgeworth expansion agrees with the formal expansion obtained by the delta method.

The three interpretations of expansions associated with $\sqrt{n}(T_n - \theta)$ will also apply to expansions associated with $\sqrt{n}(T_n' - \theta)$.

2.8. Problems. Theorem 2.1, along with its application and extensions, suggests a variety of interesting problems.

1. In the case of curved exponentials (with absolutely continuous dominating measure),

$$\sup_B \left| P_{\theta_0}\left\{\sqrt{n}\left(\hat\theta - \theta_0\right) \in B\right\} - \int_B \psi_{s;n}(y)\,dy \right| = o(n^{-(s-2)/2}).$$

   Is this true in the general case (under suitable regularity conditions and absolutely continuous dominating measure)?
2. Weakening of assumptions:
   (A) Suppose $\sum l_i z_1^{(i)}$ satisfies Condition D (or the weaker condition of Cramér). Then is the conclusion of Theorem 2.1 true at least for $s = 4$ (possibly with stronger moment assumptions)?
   (B) How far can one weaken the moment assumption? The best results seem to be those of Babu and Bai (1992).

3. Validity of formal Edgeworth expansion. Suppose

$$p = \phi_\Sigma\left(1 + \sum_1^{s-2} \frac{A_i(Z)}{(\sqrt{n})^i}\right)$$

and

$$W_n = B_1(Z) + \frac{B_2(Z)}{\sqrt{n}} + \cdots + \frac{B_{s-1}(Z)}{(\sqrt{n})^{s-2}},$$

where A's and B's are polynomials free of n, and ϕ_Σ is the multivariate normal density with zero mean and dispersion matrix Σ. The c.f. of W_n under p will have an expansion in powers of $n^{-1/2}$. Suppose the coefficient of $n^{-j/2}$ $(j = 0, 1, \ldots, n-2)$ is the c.f. of a (signed) density q_j. Is it then always true that

$$\int_{W_n < t} p\, d\underset{\sim}{Z} = \sum_{j=0}^{s-2} \frac{1}{(\sqrt{n})^j} \int_{-\infty}^{t} q_j(y)\, dy + o(n^{-(s-2)/2})? \tag{2.23}$$

Theorems in Bhattacharya and Ghosh (1978) and Chandra and Ghosh (1979) may be thought of as examples where (2.23) holds. Bhattacharya (personal communication) has pointed out that (2.23) does not hold uniformly in t without additional assumptions: Let p be standard one-dimensional normal, $W_n = Z^2/2 + 1/n^{1/2}$. If $t = 1/n^{1/2}$, then the left-hand side of (2.23) is zero, while the first term on the right is $O(n^{-1/4})$.