1. INTRODUCTION

1.1. MOTIVATION

Suppose we are interested in the density f_n of some statistic T_n $(x_1, \cdots, x_n)$, where $x_1, \cdots, x_n$ are n independent identically distributed (iid) observations with the underlying density f. Unless T_n and/or f have special forms, one cannot usually compute analytically the distribution of T_n.

A first alternative is to rely on asymptotic theory. Very often one can "linearize" the statistic T_n and prove that the linearized statistic is equivalent to T_n as $n \to \infty$, that is the difference goes to zero in probability. This leads through the central limit theorem to many asymptotic normality proofs and the resulting asymptotic distribution can be used as an approximation to the exact distribution of T_n. This is certainly a powerful tool from a theoretical point of view as can be seen in some good books on the subject, e.g. Bhattacharya and Rao (1976), Serfling (1980); cf. also the innovative article by Pollard (1985). But, in spite of the fact that in some complex situations one does not have any viable alternatives, very often the asymptotic distribution does not provide a good approximation unless the sample size is (very) large. Moreover, these approximations tend to be inaccurate in the tails of the distribution.

Many techniques have been devised to increase the accuracy of the approximation of the exact density f_n. A well known method is to use the first few terms of an *Edgeworth expansion* (cf. for instance Feller, 1971, Chapter 16). This is an expansion in powers of $n^{-1/2}$, where the leading term is the normal density. It turns out in general that the Edgeworth expansion provides a good approximation in the center of the density, but can be inaccurate in the tails where it can even become negative. Thus, the Edgeworth expansion can be unreliable for calculating tail probabilities (the values usually of interest) when the sample size is moderate to small.

In a pioneering paper, H.E. Daniels in 1954 introduced a new type of idea into statistics by applying *saddlepoint techniques* to derive a very accurate approximation to the distribution of the arithmetic mean of $x_1, \cdots, x_n$. The key idea is as follows. The density f_n can be written as an integral on the complex plane by means of a Fourier transform. Since the integrand is of the form $\exp(nw(z))$, the major contribution to this integral for large n will come from a neighborhood of the saddlepoint z_0, a zero of $w'(z)$. By means of the *method of steepest descent*, one can then derive a complete expansion for f_n with terms in powers of n^{-1}. Daniels (1954) also showed that this expansion is the same as that obtained using the idea of the *conjugate density* (see Esscher, 1932; Cramér, 1938; Khinchin, 1949) which can be summarized as follows. First, recenter the original underlying distribution f at the point t where f_n is to be approximated; that is, define the conjugate (or associate) density of f, h_t. Then use the Edgeworth expansion locally at t with respect to h_t and transform the results back in terms of the original density f. Since t is the mean of the conjugate density h_t, the Edgeworth expansion at t with respect to h_t is in fact an expansion in powers of n^{-1} and provides a good approximation locally at that point. Roughly speaking, a higher order approximation around the center of the distribution is replaced by local low order approximations around each point. The unusual characteristic of these expansions is that the first few terms (or even just the leading term) often give very accurate approximations in the far tails of the distribution even for very small sample sizes. Besides the theoretical reasons, one empirical reason for the excellent small sample behaviour is that saddlepoint approximations are density-like objects and do not show the polynomial-like waves exhibited for instance by Edgeworth approximations.

Another approximation closely related to the saddlepoint approximation was introduced independently by F. Hampel in 1973 who aptly coined the expression *small sample asymptotics* to indicate the spirit of these techniques. His approach is based on the idea of recentering the original distribution combined with the expansion of the logarithmic derivative f_n'/f_n rather than the density f_n itself. Hampel argues convincingly that this is the simplest and most natural quantity to expand. A side result of this is that the normalizing constant — that is, the constant that makes the total mass equal to 1 — must be determined numerically. This proves to be an advantage since this rescaling improves the approximation. In some cases it can even be showed that the renormalization catches the term of order n^{-1} leaving the approximation with a relative error of order $0(n^{-3/2})$; cf. Remark 3.2, section 3.3.

The aim of this monograph is to give an introduction into concepts, theory, and applications of small sample asymptotic techniques. As the title suggests, we want to include under this heading all those techniques which are similar in the spirit to those sketched above. To be a little extreme, we want to consider "*asymptotic* techniques which work well for $n = 1$" as it has been sometimes asked from a good asymptotic theory. A very simple example in this direction is Stirling's approximation to $n!$. Exhibit 1.1 shows that the *relative error* of Stirling's approximation is never greater than 4% even down to $n = 2$.

n	$n!$	Stirling approx.	relative error (%)
1	1	0.92	8.0
2	2	1.92	14.0
3	6	5.84	2.7
4	24	23.51	2.0
5	120	118.02	1.6

Exhibit 1.1

Stirling approximation ($= \sqrt{2\pi n}(n/e)^n$) to $n!$ and relative error = | exact — approx. | / exact in %.

Note that Stirling's formula is just the leading term of a Laplacian expansion of the gamma integral defining $n!$. The original approximation, that is

$$\sqrt{2\pi}((n + \frac{1}{2})/e)^{n+\frac{1}{2}},$$

derived from

$$\Delta \log x! \sim \frac{d}{dx} \log(x + \frac{1}{2})!$$

is even more accurate, cf. Daniels (1955).

Both authors were introduced into the topic via the paper by Hampel (1973) whose original idea was motivated by the application of these techniques in robust statistics. In fact, since robust procedures are constructed to be stable in a neighborhood of a fixed statistical model, their distribution theory is more complicated than that of classical procedures like least squares estimators and F-test evaluated at the normal model. In particular, it is almost impossible to compute the exact distribution of robust procedures for a finite sample size. On the other hand, the approximations based on the asymptotic distribution are often too crude to be used in practical statistical analysis. Thus, small sample asymptotics offer the tools to compute good finite sample approximations for densities, tail areas, and confidence regions based on robust statistics. The scope of there approximations is quite broad, and they have been successfully used for likelihood and conditional inference and nonparametric statistics in addition to robust statistics. Examples and computations are provided in the later chapters.

1.2. OUTLINE

The monograph is organized as follows.

In chapter 2 we review briefly the Edgeworth expansions. Although this monograph does not focus directly on Edgeworth expansions, they nevertheless play an important role as local approximations in small sample asymptotics. Therefore, we do not claim to cover the large amount of literature in this area but we just review in this chapter the basic results which will be used in the development of small sample asymptotic techniques. A reader who is aleady familiar with Edgeworth expansions can skip this chapter and go directly to chapter 3.

Chapter 3 introduces the basic idea behind saddlepoint approximations from two different points of view, namely through the method of steepest descent and integration on the complex plane (sections 3.2 and 3.3) and through the method of conjugate distributions (section 3.4). The technique is derived for a simple problem, namely the approximation of the distribution of the mean of n iid random variables.

Chapter 4 shows that small sample asymptotic techniques are available for general statistics. In particular, we discuss the approximation of the distribution of L-estimators (section 4.4) and multivariate M-estimators (section 4.5) for an arbitrary underlying distribution of the observations. In each case the theoretical development is accompanied by numerical examples which show the great accuracy of these approximations.

Chapter 5 emphasizes the relationship among a number of related techniques. First, Hampel's approach is discussed in detail in section 5.2. The relation between small sample asymptotics and large deviations is presented in section 5.3. Moreover, we attempt to relate the work by Durbin and Barndorff-Nielsen (see the review paper by Reid (1988) and the references thereof) in the case of sufficient and/or exponential families to the techniques discussed so far.

In chapter 6 we present tail areas approximations and the computation of confidence intervals for multiparameter problems, especially regression. A connection to the bootstrap and a nonparametric version of small sample asymptotics obtained by replacing the underlying distribution with the empirical distribution are also discussed.

Finally, chapter 7 is devoted to some miscellaneous aspects. In section 7.1 we discuss the computational issues. In fact, it is the availability of cheap computing which makes feasible the use of small sample asymptotic techniques for complex problems. A low order, simple approximation requiring non-trivial computations is carried out at a number of points and this is the type of problem ideally suited to computers. Section 7.2 presents a potential application of small sample asymptotics as a smoothing procedure leading to nonparametric density estimation. Some applications to robust statistics are developed in section 7.3. Finally, in the remaining sections we discuss the applications of these techniques to several different problems, including the considerable amount of work in the engineering literature.