7. FURTHER APPLICATIONS

7.1. COMPUTATIONAL ASPECTS

To evaluate the small sample approximations developed in the previous chapters we require the use of numerical software. The most critical aspect involves the solution for the saddlepoint, $\alpha(\mathbf{t})$. In the most general case, this involves solving the system

$$\int \psi_j(x,\mathbf{t}) \exp\left\{\sum_{i=1}^{p} \alpha_j(\mathbf{t})\psi_j(x,\mathbf{t})\right\} f(x)dx = 0 \quad j = 1,\cdots,p. \tag{7.1}$$

This is a non-linear system of equations which require an integration to evaluate the left hand side. Given the values for α, the computation of $c(\mathbf{t})$, $\Sigma(\mathbf{t})$, $A(\mathbf{t})$ also involve the evaluation of integrals. In all the examples for which we have done computations, x has been univariate. In principle, there is no difficulty with multivariate x, but the problems of integration become severe as the dimension of x increases.

In terms of computational effort, it is the solution of (7.1) which is the dominant feature. In the problems involving location/scale, the functions involved in (7.1) are continuous and piecewise differentiable and the resulting $\alpha_j(\mathbf{t})$'s have been smooth. This smoothness is very helpful in solving (7.1) over a grid of t since we are able to get good initial guesses for α based on values at adjacent grid points. For one dimensional problems, we have used the NAG (Numerical Algorithm Group) subroutine C05AJF. This procedure works well given a good initial guess and iterates using a secant method. The algorithm solves a sequence of problems $h(\alpha) - \theta_r h(\alpha_0)$ where $1 = \theta_0 > \theta_1 > \theta_2 \cdots > \theta_m = 0$ where α_0 is the initial guess. For each θ_r, a robust secant iteration is used based on the solution from earlier problems. If bounds are available on the solution α, a routine such as C05ADF may be used. In the case of location/scale problems α is two-dimensional. For the problem of Huber's Proposal 2, we used the NAG subroutines C05NBF or C05NCF. Both routines are suitable for the situation where the derivatives with respect to α are not provided. These routines are similar to the IMSL routines HYBRD and HYBRD1. It is necessary to choose initial estimates carefully using information from other grid points. It is possible to evaluate the derivatives of α and use a more reliable routine. However the derivatives are somewhat complicated and it is not clear that the extra coding required is worth the effort.

The numerical integration has been carried out using Gaussian quadrature. The particular NAG subroutine used is D01FBF. From experience, it appears that a 32 point quadrature procedure is necessary to get reliable results in (7.1) for t in the tail of the distribution.

The evaluation of tail areas has been carried out using the Lugannani-Rice approximation (6.3). The error in that approximation is about the same order as that resulting from a numerical integration of $f_n(\mathbf{t})$.

To give a sense of the computing time involved, we did several runs on a SUN4, 0S (4.2 Berkeley Unix) using the NAG subroutines mentioned above. The first two cases involved computation of $\alpha(t)$, $c(t)$, $s(t)$ for the mean, $\psi(x,t) = (x-t)$ for a uniform density on $[-1,1]$ and an extreme density. The results are as follows.

Uniform:	$t = 0[.005].99$	CPU time = 2.5 seconds
Extreme:	$t = -9[.1]2.5$	CPU time = 5.9 seconds

In a two dimensional problem, we solved for $\alpha_1(t)$ and $\alpha_2(t)$ in the case of the extreme density. For a total of 110 grid points (t_1 varying, t_2 fixed), the CPU time was 8.2 seconds.

The root finding procedure seems to fail if the initial values of a are not close to the final solution.

Our only attempt to find α in a three-dimensional problem was not successful. The root finding subroutine was unable to find a solution within a reasonable numbers of function evaluations. It is likely that tailor-made root finders would have to be constructed for these higher dimension problems.

It is possible to write equation (7.1) as a differential equation and then use a differential equation solver to obtain $\alpha(t)$. For the case $p = 1$, we can write (7.1) as

$$\int (\psi'(x,t) + \alpha'\psi^2(x,t) + \alpha(t)\psi(x,t)\psi'(x,t)) \exp\{\alpha(t)\psi(x,t)\} f(x)dx = 0$$

where $'$ represents differentiation with respect to t. Viewing this as a differential equation in $\alpha(t)$, we are able to solve for $\alpha(t)$. Some preliminary runs gave results similar in terms of accuracy and time to those obtained via the rootfinder. This approach seems worth pursuing in higher dimensional problems.

7.2. EMPIRICAL SMALL SAMPLE ASYMPTOTICS AND DENSITY ESTIMATION

The small sample asymptotic approximation of the density or the tail area of some statistic developed in chapters 3 through 5 requires the knowledge of the underlying distribution of the observations. From (4.25), for instance, we see that the underlying distribution F enters in the approximation only through the integrals defining C, A, Σ and α. To make the technique nonparametric, it is natural to consider replacing F by the empirical distribution function: we obtain then an *empirical small sample asymptotic approximation*. Notice that in this approximation the integrals which appear for instance in (4.25) become sums. This fact greatly reduces the complexity of the computations.

Feuerverger (1989) studied this approximation for the univariate mean and showed that an appropriate standardization of the estimator allows to keep a relative error of order $O_p(n^{-1/2})$. Ronchetti and Welsh (1990) extended this result to multivariate M-estimators and proved that by appropriate standardization of the estimator the renormalized empirical approximation has a relative error or order $O_p(n^{-1})$. In section 6.3 a similar idea was used to construct confidence intervals based on the empirical cumulant generating function. In that situation, the exponential tilt was used to adjust the first cumulant enabling us to obtain second order coverage for the confidence interval. To summarize: the empirical small sample asymptotic approxiamtion is an alternative to the bootstrap with the advantage that resampling is avoided. For a related result, see Davison and Hinkley (1988).

A related potential application is the use of small sample asymptotics in density estimation. The approximations which have been developed are asymptotic expansions in terms of n. Since they work so well for very small sample sizes, we can examine them with $n = 1$ although clearly it does not make sense to discuss the order of the error terms. The approximation for $n = 1$ can be thought of as a version of the underlying density smoothed in some sense towards normality.

Consider the small sample asymptotic approximation of univariate M-estimators given by (4.8). For $n = 1$ we obtain

$$g_1(t) \propto A(t)/(c(t)\sigma(t)),$$

where A, c, and σ are defined in Theorem 4.3. $g_1(t)$ depends on the function ψ which defines the M-estimator. Let us consider this approximation for the mean ($\psi(x,t) = x - t$) and the

following four underlying situations:

1. Uniform on $[-1, 1]$
2. Extreme density, $f(x) = \exp(x - \exp(x))$
3. Cosine density on $[0, 1]$, $f(x) = 1 + \cos 4\pi x$
4. U-shape density on $[-1, 1]$, $f(x) = 1.5x^2$

The last two densities were chosen as situations where the density is very different from a normal and we might expect the approximation to do badly. For each of the above, we have calculated g_1 and plotted it with the true underlying density f in Exhibit 7.1. As can be seen from the plots, g_1 is a reasonable estimate of the density for the extreme but as expected is a poor estimate for the cosine and U-shape although it starts to match reasonably well in the tails. Exhibits 3.9 through 3.13 confirm this result.

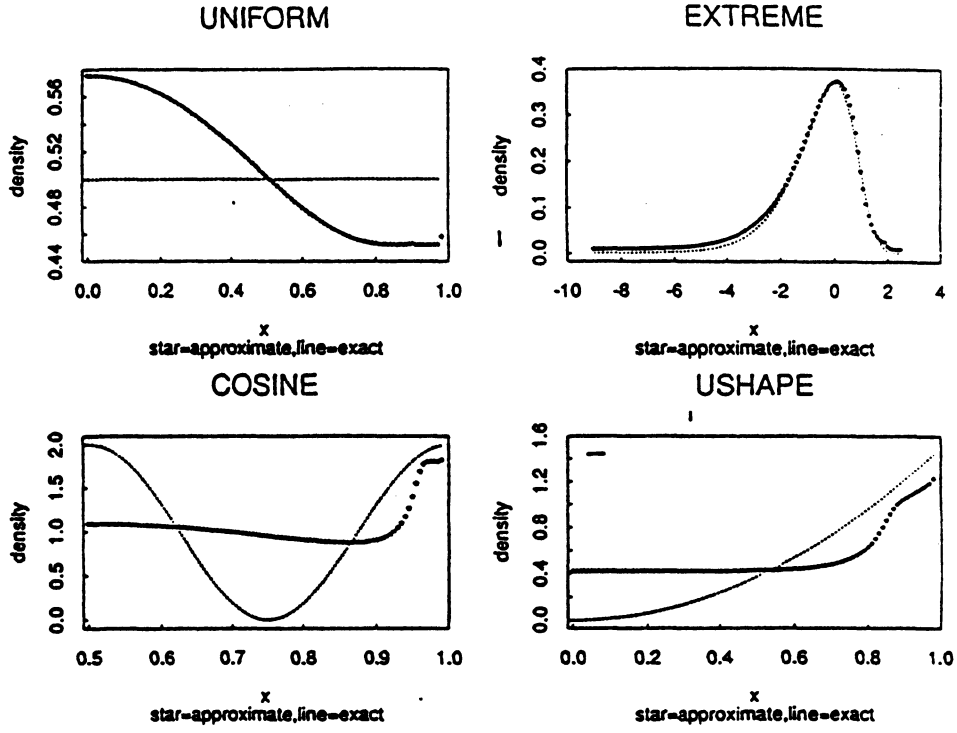

Exhibit 7.1
Small sample approximation with $n = 1$ versus true density

If we want to use g_1 as a density estimator, we face two problems. Firstly, we have different possible choices of ψ. Each of them will define a different estimator. Secondly, the underlying density (which we are trying to estimate) enters in the computation of g_1 through A, c, and σ.

The second problem can be solved by replacing the underlying distribution in A, c, and σ by the empirical distribution function. If we apply this idea to $g_1(t)$ we obtain the following *nonparametric density estimator* $\hat{g}_1(t)$ for $f(t)$ (see Ronchetti, 1989):

$$\hat{g}_1(t) = \hat{D}_1 \hat{A}(t)/(\hat{c}(t)\hat{\sigma}(t)), \tag{7.2}$$

where $\hat{\alpha}(t)$ is determined by the implicit equation

$$\sum_{i=1}^{n} \psi(x_i, t) \exp\{\hat{\alpha}(t)\psi(x_i, t)\} = 0,$$

and

$$\hat{c}(t) = \left(\frac{1}{n}\sum_{i=1}^{n}\exp\{\hat{\alpha}(t)\psi(x_i,t)\}\right)^{-1},$$

$$\hat{A}(t) = \hat{c}(t)\frac{1}{n}\sum_{i=1}^{n}\frac{\partial\psi(x_i,t)}{\partial t}\exp\{\hat{\alpha}(t)\psi(x_i,t)\},$$

$$\hat{\sigma}^2(t) = \hat{c}(t)\frac{1}{n}\sum_{i=1}^{n}\psi^2(x_i,t)\exp\{\hat{\alpha}(t)\psi(x_i,t)\},$$

$$\hat{D}_1 = \left(\int A(t)/(\hat{c}(t)\hat{\sigma}(t))dt\right)^{-1}.$$

We can investigate the quality of this estimator by looking at $\hat{g}_1(t) - f(t)$. This difference can be written as

$$\hat{g}_1(t) - f(t) = [\hat{g}_1(t) - g_1(t)] + [g_1(t) - f(t)].$$

Whereas the first term (variability) decreases as n increases, the second one (bias) is fixed. Clearly this term depends on the choice of ψ and on the underlying density f and plays an important role in determining the quality of the estimator.

Feuerverger (1989) studied this problem in the case of the mean ($\psi(x,t) = x - t$) and showed that the bias can be substantial; see Exhibit 7.2.

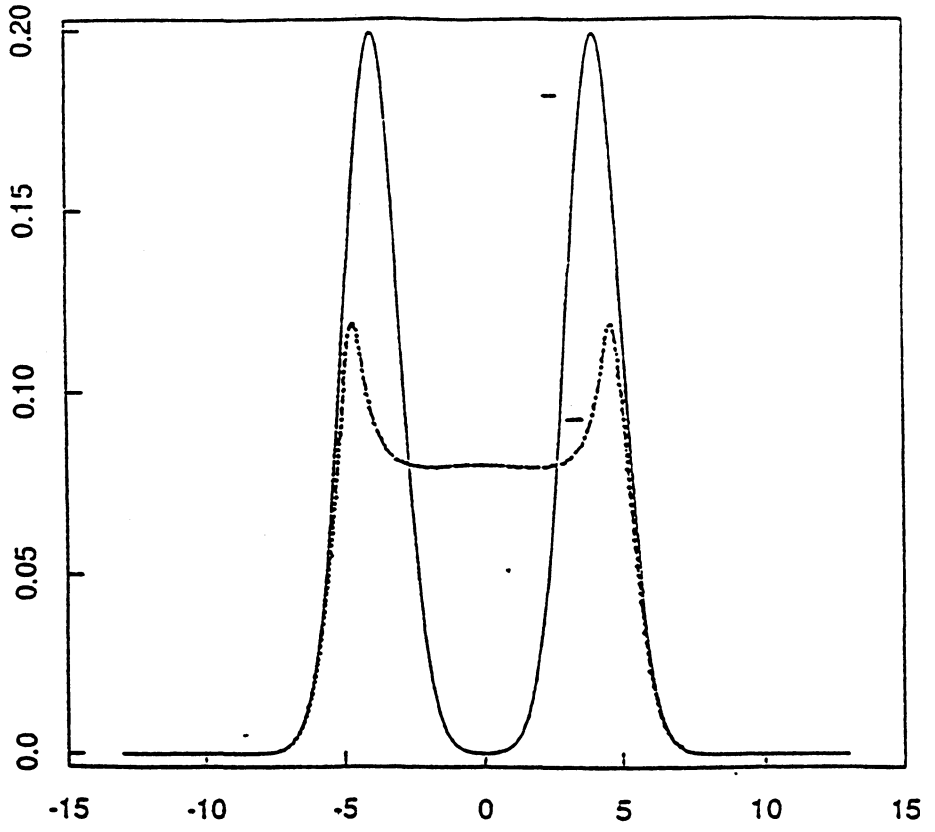

Exhibit 7.2
$f(t)$ and $g_1(t)$ (- - -) for the underlying distribution
$.5^*N(4,1) + .5^*N(-4,1)$; from Feuerverger (1989).

Ronchetti (1989) proposed an estimator based on the following ψ function:

$$\begin{aligned} \psi_{k_n}(y) &= y/k_n \qquad \text{if } |y| \le k_n \\ &= sgn(y) \qquad \text{otherwise,} \end{aligned}$$

where $k_n \to 0$ as $n \to \infty$. With this choice of ψ, (7.2) takes the form:

$$\hat{g}_1(t) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{2k_n} 1\left\{\frac{|x_i - t|}{k_n} \le 1\right\} \exp\{\alpha_n(t)\psi_{k_n}(x_i - t)\} \cdot \left[\frac{\sum_j \exp\{\alpha_n(t)\psi_{k_n}(x_j - t)\}}{\sum_j \psi^2_{k_n}(x_j - t)\exp\{\alpha_n(t)\psi_{k_n}(x_j - t)\}}\right]^{1/2} \cdot \tag{7.2a}$$

where $1\{\ \}$ is the indicator function and $\alpha_n(t)$ is defined through the implicit equation

$$\sum_{i=1}^{n} \psi_{k_n}(x_i - t)\exp\{\alpha_n(t)\psi_{k_n}(x_i - t)\} = 0.$$

Note that $\hat{D}_1$ in (7.2) is replaced here by its asymptotic value 1/2. The reason for this choice of ψ is that ψ_{k_n} converges to the ψ-function of the median as $n \to \infty$ and the small sample asymptotic approximation for the median is exact (after renormalization) for any underlying density f; see Field and Hampel (1982) and section 4.2. Therefore the bias term of the density estimator (7.2*a*) vanishes asymptotically. A comparison of Exhibits 7.2 and 7.3 shows the important reduction of bias achieved by using the estimator (7.2*a*). More details are provided in Ronchetti (1989).

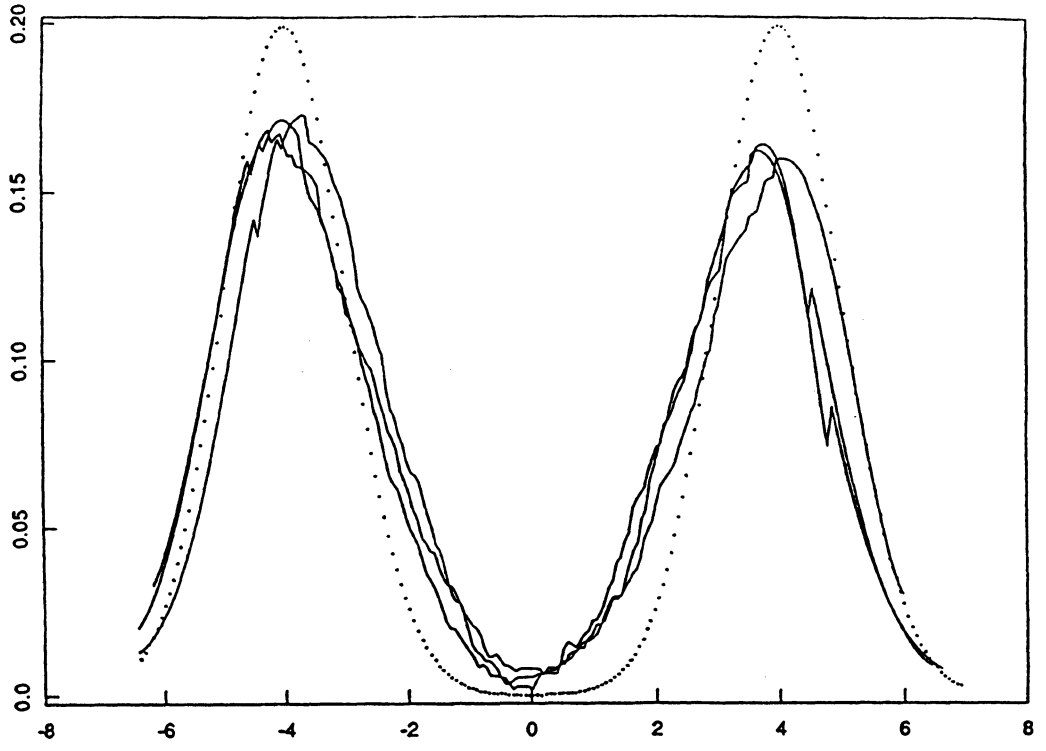

Exhibit 7.3
Exact density (- - -) and density estimator for three samples generated from a $.5^*N(4,1) + .5^*N(-4,1)$ and $n = 200$.

Finally it is possible to use $\hat{g}_1(t)$ along with the Lugannani-Rice tail area approximation to construct prediction intervals. Field and Manchester are currently carrying out some research in this direction.

7.3. ROBUST TESTING

7.3.a The Use of Small Sample Asymptotics in Robust Statistics

In this section we will discuss an application of small sample asymptotics to robust statistics. It shows that small sample asymptotics can be applied successfully not only to compute very accurate approximations to the exact distribution of robust estimators and test statistics (cf. sections 4.2 and 4.5 and Visek, 1983, 1986) but also, from a more

methodological point of view, to define new tools and to improve those based on asymptotic normality.

Consider, for instance, the study of the robustness properties of tests. Two different approaches are available.

The first one is Huber's minimax approach (see Huber 1965; 1981, p. 253 ff.) which is based on the following idea. In a simple hypothesis against a simple alternative testing problem, find the test which maximizes the minimum power over a neighborhood of the alternative, under the side condition that the maximum level over a neighborhood of the hypothesis is bounded. The solution to this problem is the censored likelihood ratio test which is based on a truncated likelihood ratio statistic. This ensures that outlying points will have only a bounded influence on the level and on the power of the test. While this formulation of the problem is very elegant and leads to an exact finite sample result, it seems very difficult to generalize it beyond the simple situation described above.

The second approach is based on the idea of influence function originally introduced by Hampel (1968, 1974) for estimators; see section 2.5. Here the key point is to investigate the behaviour of the level and of the power of a test when the true underlying distribution of the observations doesn't belong to the model F_θ but is of the form $(1-\epsilon)F_\theta+\epsilon\Delta_x$ ("contaminated distribution"), where $0 < \epsilon < 1$, and Δ_x is the distribution which puts mass 1 at the point x. The actual level and the actual power computed under the contaminated distribution will describe, as functions of x, the influence on the level and on the power of a small amount of contamination at the point x. A natural robustness requirement for these functions is to be bounded in x. For details on this approach we refer to Ronchetti (1979, 1982, 1987), Rousseeuw and Ronchetti (1979, 1981) and Hampel et al. (1986), chapters 3 and 7. A related technique developed by Lambert (1981) is to look at the log P-value of the test rather than the level and the power. The corresponding influence function is related to the previous one as shown in Hampel et al. (1986), section 3.6. While the influence function approach is very general and can be applied to complex situations, as for instance testing in the regression model (cf. Ronchetti, 1982, 1987, and Hampel et al., 1986, chapter 7), it requires the evaluation of tail areas under contaminated distibutions. Since in general these cannot be computed exactly, one has to approximate them by means of the asymptotic distribution of the test statistic.

Small sample asymptotic techniques which lead to very accurate approximations down to very small sample sizes can be used to bridge the gap between exact finite sample results and results based on the asymptotic distribution. In sections 7.3b and 7.3c we will illustrate this point for a simple class of tests. The arguments hold for more general situations.

7.3.b A Tail Area Influence Function

Suppose we want to investigate the properties of the tail area $P_F(T_n > t)$ (t and n fixed) from a robustness point of view or, more precisely, we are interested in the behavior of $P_F(T_n > t)$ when the observations do not follow exactly the model distribution F. To accomplish this, we use the idea of an influence function and introduce a tail area influence function.

Definition 7.1 Let Δ_x be the distribution which puts mass 1 at any point $x \in \mathbf{R}$ and define $F_{\epsilon,x} := (1-\epsilon)F + \epsilon\Delta_x$. Then, the *tail area influence function* of T_n at F is defined by

$$TAIF(x;t;T_n,F) = \lim_{\epsilon\to 0}\Big[P_{F_{\epsilon,x}}(T_n > t) - P_F(T_n > t)\Big]/\epsilon,$$

for all $x \in \mathbf{R}$ where the right hand side exists.

$TAIF(x;t;T_n,F)$ describes the normalized influence on the tail area of a small amount of contamination at the point x.

Let us now apply this definition to the following one-parameter situation. Consider a family of distributions $\{F_\theta|\theta \in \Theta\}$ where Θ is a real interval and a sequence of statistics $T_n(x_1,\cdots,x_n)$ which are used to test the null hypothesis $H_0 : \theta = \theta_0$. The test is assumed to be based on a M-statistic, that is the test statistic T_n is the solution of the equation

$$\sum_{i=1}^{n}\psi(x_i,T_n) = 0, \tag{7.3}$$

where $x_1,\cdots,x_n$ are n independent observations and ψ is a given function; cf. (4.1).

A small sample asymptotic approximation to the exact density $f_n(t)$ can be obtained from Theorem 4.3. The corresponding approximation to the tail area $P_F(T_n > t) = \int_t^\infty f_n(u)du$ can be computed by means of a Laplacian approximation to the integral when the approximation (4.8) is substituted for $f_n(u)$; cf. section 6.2. This leads to the following tail area approximation

$$P_F(T_n > t) = (2\pi n)^{-1/2}c_F^{-n}(t)/\big[\sigma_F(t)\cdot\alpha_F(t)\big]\cdot\big[1+0(1/n)\big], \tag{7.4}$$

where $\alpha_F(t)$ solves

$$\int\psi(x,t)\exp\{\alpha_F\cdot\psi(x,t)\}dF(x) = 0, \tag{7.5}$$

$$c_F^{-1}(t) = \int\exp\{\alpha_F\cdot\psi(x,t)\}dF(x),$$

$$\sigma_F^2(t) = c_F(t)\int\psi^2(x,t)\{\alpha_F\cdot\psi(x,t)\}dF(x).$$

Notice that this approximation works as well as (6.3) in the tails; see section 6.2.

We now apply Definition 7.1 to (7.4) and compute the tail area influence function. To achieve this we have to evaluate the tail area under the contaminated distribution $F_{\epsilon,x} = (1-\epsilon)F+\epsilon\Delta_x$. Although this distribution formally does not satisfy the regularity conditions required for (7.4) to hold (see sections 4.2 and 6.2), one can first consider a mixture $(1-\epsilon)F+\epsilon G_\delta$, where G_δ has a density and the property that as δ approaches to 0, G_δ approaches Δ_x. Then, (7.4) holds for the mixture and a limiting argument can be used to show that it holds for $F_{\epsilon,x}$. Using the defining equation (7.5) for α we obtain

$$(\partial/\partial\epsilon)c_{F_{\epsilon,x}}^{-n}(t)\big|_{\epsilon=0} = nc_F^{-n}(t)\cdot\big[c_F(t)\cdot\exp(\alpha_F(t)\cdot\psi(x,t)) - 1\big]$$

and

$$\begin{aligned}(\partial/\partial\epsilon)&\Big[c_{F_{\epsilon,x}}^{-n}(t)/\big(\sigma_{F_{\epsilon,x}}(t)\cdot\alpha_{F_{\epsilon,x}}(t)\big)\Big]_{\epsilon=0}\\ &= \big\{c_F^{-n}(t)/\big[\sigma_F(t)\cdot\alpha_F(t)\big]\big\}\cdot\big\{n\big[c_F(t)\cdot\exp(\alpha_F(t)\cdot\psi(x,t)) - 1\big]\\ &\quad - (\dot\alpha_F/\alpha_F)(t) - (\dot\sigma_F/\sigma_F)(t)\big\},\end{aligned}$$

where the dot denotes the derivative with respect to ϵ at $\epsilon = 0$. Therefore, the tail area influence function of a M-statistic defined by a function ψ (see (7.3)) is given by

$$TAIF(x;t;\psi,F) = (2\pi)^{-1/2}\{c_F^{-n}(t)/[\sigma_F(t)\cdot\alpha_F(t)]\}\cdot$$
$$\{n^{1/2}[c_F(t)\cdot\exp(\alpha_F(t)\cdot\psi(x,t)) - 1] + 0(n^{-1/2})\}. \tag{7.6}$$

The $TAIF$ can be interpreted in the same way as any influence function; see Hampel et al. (1986), chapters 2, 3. For instance, if $K_n^{(a)}$ is the critical value of a test of nominal level a based on (7.3), $TAIF\ (x;K_n^{(a)};\psi,F_{\theta_0})$ is the influence function of the level of this test and describes the robustness of validity of the test. A bounded (in x) $TAIF\ (x;K_n^{(a)};\psi,F_{\theta_0})$ indicates that the maximum influence of a single observation on the level of the test is bounded. Hence, the test has robustness of validity. Similarly for the power and robustness of efficiency.

In other words, $TAIF$ when applied to a testing problem, can be viewed as a small sample refinement of the (asymptotic) level influence function (LIF) and the (asymptotic) power influence function (PIF) discussed in Hampel et al. (1986), chapter 3. In fact the following theorem holds.

Theorem 7.1

Given $0 < a < 1$ (level), let $K_n^{(a)}$ (critical value) be defined by $P_{F_{\theta_0}}(T_n > K_n^{(a)}) = a$. Define a sequence θ_n of alternatives as $\theta_n = \theta_0 + \delta n^{-1/2}$, $\delta > 0$.

Then, under the assumptions A4.1-A4.5 of section 4.2 the following holds:

(*i*) $$\lim_{n\to\infty} n^{-1/2}TAIF(x;K_n^{(a)};\psi,F_{\theta_0})$$
$$= \phi(\Phi^{-1}(1-a))\cdot\psi(x,\theta_0)/(E\psi^2)^{1/2} = \ \mathrm{LIF}(x;\psi,F_{\theta_0}),$$

(*ii*) $$\lim_{n\to\infty} n^{-1/2}TAIF(x;K_n^{(a)};\psi,F_{\theta_n})$$
$$= \phi(\Phi^{-1}(1-a) - \delta|E\psi'|/(E\psi^2)^{1/2})\cdot\psi(x;\theta_0)/(E\psi^2)^{1/2}$$
$$= PIF(x;\psi,F_{\theta_0}),$$

where ϕ and Φ are the density and the cumulative of the standard normal distribution, respectively, and $E\psi' = \int\psi'(x,\theta_0)dF_{\theta_0}(x)$, $E\psi^2 = \int\psi^2(x,\theta_0)dF_{\theta_0}(x)$ and ψ' denotes differentiation with respect to θ.

Proof: We prove (i). More tedious computations with similar arguments lead to (ii). By (7.6) we have to compute

$$\lim_{n\to\infty}(2\pi)^{-1/2}\left[c_{F_{\theta_0}}^{-n}(\cdot)/(\sigma_{F_{\theta_0}}(\cdot)\alpha_{F_{\theta_0}}(\cdot))\right]\cdot$$
$$\left[c_{F_{\theta_0}}(\cdot)\exp(\alpha_{F_{\theta_0}}(\cdot)\cdot\psi(x,\cdot)) - 1 + 0(n^{-1})\right], \tag{7.7}$$

where the argument $(\cdot)$ of the functions involved is $K_n^{(a)}$. From the asymptotic normality of T_n (see Huber 1967, 1981)

$$L_{F_{\theta_0}}(n^{1/2}(T_n - \theta_0)) \to N(0, E\psi^2/(-E\psi')^2),$$

as $n \to \infty$, we obtain

$$K_n^{(a)} = \theta_0 + n^{-1/2}\Phi^{-1}(1-a)\cdot\left[E\psi^2/(E\psi')^2\right]^{1/2} + 0(n^{-1}). \tag{7.8}$$

Using the defining equations for c, α, σ (see (7.5)) and the Fisher-consistency of T, that is $T(F_{\theta_0}) = \theta_0$, we get

$$K_n^{(a)} \to \theta_0,\ \alpha_{F_{\theta_0}}(K_n^{(a)}) \to 0,\ c_{F_{\theta_0}}(K_n^{(a)}) \to 1,\ \sigma^2_{F_{\theta_0}}(K_n^{(a)}) \to E\psi^2,$$

when $n \to \infty$.

Now define $B_n := \log c^{-n}_{F_{\theta_0}}(K_n^{(a)})$.

$$\lim_{n\to\infty} B_n = -\lim_{n\to\infty}\left[\log c_{F_{\theta_0}}(K_n^{(a)})\right]/n^{-1}$$

and by L'Hôpital's rule and (7.8)

$$= -\frac{1}{2}\Phi^{-1}(1-a)\cdot\left[E\psi^2/(E\psi')^2\right]^{1/2}\cdot\lim_{n\to\infty}(c'_{F_{\theta_0}}/c_{F_{\theta_0}})(K_n^{(a)})/n^{-1/2}, \tag{7.9}$$

where $c'_{F_{\theta_0}}(t) = (\partial/\partial t)c_{F_{\theta_0}}(t)$. By the defining equation (7.5) we have

$$c'_{F_{\theta_0}}(t) = -c_{F_{\theta_0}}(t)\cdot\alpha_{F_{\theta_0}}(t)\cdot A_{F_{\theta_0}}(t),$$

where $A_F(t) = c_F(t)\int\psi'(x,t)\exp\{\alpha_F\cdot\psi(x,t)\}dF(x)$, and $A_{F_{\theta_0}}(K_n^{(a)}) \to E\psi'$, as $n \to \infty$. Therefore, from (7.8) we obtain

$$\lim_{n\to\infty} B_n = \frac{1}{2}\Phi^{-1}(1-a)\cdot(E\psi^2)^{1/2}sgn(E\psi')\cdot\lim_{n\to\infty}\alpha_{F_{\theta_0}}(K_n^{(a)})/n^{-1/2}. \tag{7.10}$$

Since (dropping the arguments)

$$\alpha' = -\left[\int\psi'\exp(\alpha\cdot\psi)dF + \alpha\cdot\int\psi\cdot\psi'\exp(\alpha\cdot\psi)dF\right]/\int\psi^2\cdot\exp(\alpha\cdot\psi)dF,$$

using L'Hôpital's rule in (7.10), we have

$$\lim_{n\to\infty} B_n = -\frac{1}{2}\left[\Phi^{-1}(1-a)\right]^2, \tag{7.11}$$

and finally

$$\lim_{n\to\infty} c^{-n}_{F_{\theta_0}}(K_n^{(a)}) = (2\pi)^{1/2}\phi(\Phi^{-1}(1-a)).$$

Therefore,

$$\lim_{n\to\infty} n^{-1/2}TAIF(x; K_n^{(a)}; \psi, F_{\theta_0})$$

$$= \phi(\Phi^{-1}(1-a))/(E\psi^2)^{1/2}\cdot\lim_{n\to\infty}\left[c_{F_{\theta_0}}(\cdot)\cdot\exp(\alpha_{F_{\theta_0}}(\cdot)\cdot\psi(x,\cdot)) - 1\right]/\alpha_{F_{\theta_0}}(\cdot)$$

$$= \left[\phi(\Phi^{-1}(1-a))\right]/(E\psi^2)^{1/2}\cdot\psi(x,\theta_0),$$

and this equals $LIF(x;\psi,F_{\theta_0})$ as defined in Hampel et al. (1986), (3.2.13), p. 199 and p. 204.

□

7.3.c Approximation of The Maximum Level and Minimum Power Over Gross-error Models

Influence functions can be used to extrapolate the value of the functional of interest over a neighborhood of a given model distribution F; cf. Hampel et al. (1986), p. 173, 175, 200. Consider, for instance, the gross-error model

$$P_\epsilon(F) := \{G : G = (1-\epsilon)F + \epsilon H, \quad H \text{ arbitrary distribution}\}.$$

Using the first two terms of the von Mises expansion (cf. von Mises 1947, Hampel et al. 1986, p. 85) we obtain

$$P_{(1-\epsilon)F+\epsilon H}(T_n > t) \cong P_F(T_n > t) + \epsilon \int TAIF(x;t;T_n,F)dH(x),$$

hence

$$s(t;T_n,F,\epsilon) := sup_H P_{(1-\epsilon)F+\epsilon H}(T_n > t) \cong P_F(T_n > t) + \epsilon \cdot sup_x TAIF(x;t;T_n,F). \tag{7.12}$$

Similarly for the infimum we get

$$i(t;T_n,F,\epsilon) := inf_H P_{(1-\epsilon)F+\epsilon H}(T_n > t) \cong P_F(T_n > t) + \epsilon \cdot inf_x TAIF(x;t;T_n,F). \tag{7.13}$$

(7.12) and (7.13) require the approximation to be valid uniformly over shrinking ϵ-contamination neighbourhoods. For M-statistics defined by (7.3), this can be obtained by imposing some additional conditions on the function ψ as in Rieder (1980), p. 114. In general these approximations will be valid only when the amount of contamination is much smaller than the breakdown point of T_n.

This type of approximation has been used successfully for other kinds of problems, for instance the approximation of the variance of an estimator over ϵ-contaminated models, and leads to very accurate results (cf. Hampel 1983; Hampel et al. 1986, p. 173).

Let us now apply (7.12) and (7.13) to the testing problem. Set $t = K_n^{(a)}$,the critical value of a level-a test, $F = F_{\theta_0}$ in (7.12) and $F = F_\theta$, $\theta \neq \theta_0$, in (7.13). Consider the class of tests based on a M-statistic T_n defined by a function ψ (see (7.3)) and approximate $TAIF$ using (7.6). To compare with asymptotic results we replace ϵ by $\epsilon_n = \epsilon \cdot n^{-1/2}$ and we define a sequence of alternatives $\theta_n = \theta_0 + \delta n^{-1/2}$. Then we have

$$\begin{aligned} \text{supremum level} &= s(K_n^{(a)};\psi,F_{\theta_0},\epsilon_n) \\ &\cong a + \epsilon \cdot b_n(K_n^{(a)},F_{\theta_0}) \cdot \{c_{F_{\theta_0}}(K_n^{(a)}) \cdot \\ &\qquad \exp[\alpha_{F_{\theta_0}}(K_n^{(a)}) \cdot sup_x \psi(x,K_n^{(a)})] - 1\} \end{aligned} \tag{7.14}$$

and

$$\begin{aligned} \text{infimum power} &= i(K_n^{(a)};\psi,F_{\theta_n},\epsilon_n) \\ &\cong \beta(\theta_n) + \epsilon \cdot b_n(K_n^{(a)},F_{\theta_n}) \cdot \{c_{F_{\theta_n}}(K_n^{(a)}) \cdot \\ &\qquad \exp[\alpha_{F_{\theta_n}}(K_n^{(a)}) \cdot inf_x \psi(x,K_n^{(a)})] - 1\}, \end{aligned} \tag{7.15}$$

where $b_n(t,F) := (2\pi)^{-1/2} \cdot c_F^{-n}(t)/[\sigma_F(t) \cdot \alpha_F(t)]$ and $\beta(\theta_n)$ is the power of the test at the alternative F_{θ_n}. Since, in general the exact value of $K_n^{(a)}$ is difficult to compute, we can use the approximation given by (7.8).

According to Theorem 7.1, as $n \to \infty$, (7.14) becomes

$$\lim_{n\to\infty} s(K_n^{(a)}; \psi, F_{\theta_0}, \epsilon_n) \cong a + \epsilon \cdot sup_x LIF(x; \psi, F_{\theta_0}), \tag{7.16}$$

and (7.15),

$$\lim_{n\to\infty} i(K_n^{(a)}; \psi, F_{\theta_0}, \epsilon_n) \cong \beta_{as} + \epsilon \cdot inf_x PIF(x; \psi, F_{\theta_0}), \tag{7.17}$$

where $\beta_{as} = 1 - \Phi(\Phi^{-1}(1-a) - \delta[(E\psi')^2/E\psi^2]^{1/2})$ is the asymptotic power of the test at the model.

As an illustration we compare the exact results with the approximations (7.14) and (7.15) (and the approximations obtained from LIF and PIF based on the asymptotic distribution) in the following situation:

$$F_\theta(x) = \Phi(x-\theta),$$
$$H_0 : \theta = 0,$$
$$T_n = \text{ median, that is } \psi(x,t) = sgn(x-t).$$

We choose this case because it allows an explicit exact computation of the supremum of the level and the infimum of the power. Details of computations can be found in Field and Ronchetti (1985).

n	$K_n^{(a)}$	ϵ	0.00	.05	.01	.10
3	1.101		5.00	5.36	6.89	9.03
5	.880			5.26	6.37	7.94
7	.754			5.22	6.17	7.49
11	.609			5.19	5.98	7.08

Exhibit 7.4

Exact supremum of the level $s(K_n^{(a)}; med, \Phi, \epsilon_n)$ (in %) over $(1-\epsilon_n)\cdot\Phi(x) + \epsilon_n H(x)$ of the test based on the median ($a = 5\%$)

n	$K_n^{(a)}$ (approx.)	ϵ	0.00	.01	.05	.10
3	1.190		5.00	5.34	6.72	8.43
5	.922			5.25	6.24	7.48
7	.779			5.21	6.06	7.11
11	.622			5.18	5.90	6.79
100	.206			5.12	5.61	6.22
10000	.021			5.10	5.52	6.05
∞	0.000			5.10	5.51	6.03

Exhibit 7.5

Approximate supremum of the level $s(K_n^{(a)}; med, \Phi, \epsilon_n)$ (in %) over $(1-\epsilon_n)\Phi(x) + \epsilon_n H(x)$ of the test based on the median ($a = 5\%$) (approx. (7.14), (7.16)).

n	$K_n^{(a)}$	δ	ϵ 0.00	.01	.05	.10
3	1.101	.5	11.20	11.08	10.61	10.04
		1.0	21.62	21.41	20.54	19.48
		1.5	36.21	35.87	34.52	32.85
5	.880	.5	10.95	10.83	10.34	9.75
		1.0	20.85	20.63	19.77	18.71
		1.5	34.72	34.38	33.07	31.45
7	.754	.5	10.85	10.72	10.22	9.60
		1.0	20.53	10.31	19.43	18.36
		1.5	34.09	33.76	32.46	30.85
11	.609	.5	10.78	10.64	10.11	9.47
		1.0	20.28	20.05	19.15	18.06
		1.5	33.58	33.26	31.94	30.34

Exhibit 7.6

Exact infimum of the power $i(K_n^{(a)}; med, \phi, (\cdot - \theta_n), \epsilon_n)$ (in %) over $(1-\epsilon_n)\cdot\Phi(x-\theta_n)+\epsilon_n H(x)$ of the test based on the median ($a = 5\%$).

n	$K_n^{(a)}$ (approx.)	δ	ϵ 0.00	.01	.05	.10
3	1.190	.5	10.15	10.05	9.67	9.19
		1.0	19.26	19.08	18.38	17.49
		1.5	32.26	31.98	30.85	29.44
5	.922	.5	10.34	10.23	9.79	9.24
		1.0	19.49	19.29	18.52	17.54
		1.5	32.43	32.13	30.95	29.46
7	.779	.5	10.42	10.30	9.83	9.24
		1.0	19.59	19.38	18.56	17.53
		1.5	32.51	32.20	30.98	29.45
11	.622	.5	10.50	10.37	9.85	9.21
		1.0	19.68	19.46	18.59	17.50
		1.5	32.57	32.26	30.99	29.41
100	.206	.5	10.62	10.46	9.81	8.99
		1.0	19.82	19.57	18.54	17.26
		1.5	32.68	32.33	30.95	29.23
10000	.021	.5	10.63	10.45	9.73	8.82
		1.0	19.84	19.56	18.46	17.08
		1.5	32.69	32.33	30.90	29.10
∞	.000	.5	10.64	10.45	9.72	8.80
		1.0	19.85	19.57	18.45	17.06
		1.5	32.70	32.34	30.90	29.09

Exhibit 7.7

Approximate infimum of the power $i(K_n^{(a)}; med, \Phi(\cdot - \theta_n), \epsilon_n)$ (in %) over $(1-\epsilon_n)\Phi(x-\theta_n) + \epsilon_n H(x)$ of the test based on the median ($a = 5\%$) (approx. (7.13), (7.15)).

A comparison of Exhibit 7.4 and Exhibit 7.5 shows the general good accuracy of the approximation based on $TAIF$ down to small n ($= 3, 5$). Moreover, this approximation improves the asymptotic result based on LIF. For instance, for $\epsilon = 10\%$ and $n = 3$, the exact supremum of the level is 9.03%, the approximation based on $TAIF$ gives 8.43%, while the asymptotic normality predicts 6.03%.

At the moment the advantage of the asymptotic result lies in its greater flexibility: the expression of the "asymptotic" influence function is simpler and can be computed for general classes of tests. Therefore, optimal robust tests can be derived using the asymptotic theory and eventually refined for small sample sizes.

The results related to the infimum of the power (Exhibits 7.6 and 7.7) seem to indicate that here the accuracy of the approximation based on $TAIF$ is not as good as for the supremum of the level. The reason seems to lie in the approximation of the power $\beta(\theta)$

at the model. With a better correction term it is possible to achieve the high accuracy obtained for the supremum of the level.

For example, if we used in (7.15) the exact value for $\beta(\theta_n)$ instead of the approximation β_{as}, we would obtain for $n = 3$, $\delta = .5$ and $\epsilon = .01, .05, .10$ the following values for the given power: 11.10%, 10.72%, 10.24%.

7.3.d Robustness Versus Efficiency

Given a test based on a test statistic T_n defined by ψ, one can compute up to terms of the order $n^{-1/2}$ the tail area influence function (7.6) and discuss the behaviour of $TAIF$ with respect to x, t, n. One will strive for a $TAIF$ that is bounded (not necessarily in a symmetric way) with respect to x to limit the influence of outliers, that is continuous in x to limit grouping and rounding effects, etc. Moreover, putting $F = F_\theta$ and $t = K_n^{(a)}$, the critical value of a test, by means of $TAIF$ one can investigate the influence of outliers on the level and on the power and their behaviour over gross-error models (see subsection 7.3b).

A natural problem that arises is to try and find a balance between robustness and efficiency. We can look for a test in the class (7.3) that maximizes the power, under the condition of a bounded (with respect to x) $TAIF$ $(x; K_n^{(a)}; \psi, F_\theta)$. Using (7.4) and (7.6) to approximate the power and the tail area influence function, one can hope to find a function $\psi_{opt}(x,t)$ that satisfies a first order condition for the optimality problem. We conjecture that in the normal-location case, that is, $F_\theta(x) = \Phi(x-\theta)$, $\psi_{opt}(x,t)$ equals the Huber-function

$$\begin{aligned} \psi_k(x-t) &= x-t && \text{if } |x-t| \le k \\ &= k\ sgn(x-t) && \text{otherwise,} \end{aligned}$$

although an exact proof seems to be rather complicated.

7.4. SADDLEPOINT APPROXIMATION FOR THE WILCOXON TEST

In this section we discuss an application of the method of steepest descent to rank procedures. More precisely, we discuss the approximation of the density and tail areas of the Wilcoxon test statistic.

Given a sample of n independent observations $x_1, \cdots, x_n$, we want to test the null hypothesis H_0 that these observations have a common symmetric distribution about 0. To this end one can use the Wilcoxon test or signed-rank test which is defined by the test statistic

$$T_n = \sum_{i=1}^{n} R_i \cdot sgn(x_i), \tag{7.18}$$

where R_i is the rank of x_i according to its absolute value. If large values of T_n are significant, we are then interested in the tail probability

$$P = P[T_n \ge t_0 | H_0]. \tag{7.19}$$

Although tables of the exact cumulative distribution

$$q_k = P[0 \le T_n \le k | H_0] \tag{7.20}$$

are available, it is nevertheless interesting to investigate the performance of saddlepoint approximations in the case. Moreover, a good approximation of (7.19) has its own interest because it would allow to compute P-values for *any* given n and t_0. This is even more important in the two-sample case, where two parameters (the sample sizes of the two groups) are involved in addition to t_0.

There are different ways one can use to approximate (7.19). For instance, one could apply the general saddlepoint approximation developed in section 4.3. By using an Edgeworth expansion to approximate the cumulant generating function of T_n, one can approximate by means of the techniques of section 4.3 a saddlepoint expansion to the density or a saddlepoint expansion to the tail area. However, it turns out that in this case the exact probability generating function is known and therefore a direct saddlepoint approximation can be computed. This was done by Helstrom (1986b) and we follow here his development. Similar results can be found in Robinson et al. (1988).

Since T_n can assume only discrete values, define $p_j = P[T_n = j|H_0]$. Then, by Cauchy theorem

$$p_j = (2\pi)^{-1} \int_{\mathcal{P}} z^{-(j+1)} h_n(z) dz,$$

$$q_k = P[0 \le T_n \le k|H_0] = \sum_{j=0}^{k} p_j = (2\pi i)^{-1} \int_{\mathcal{P}} \left\{ \sum_{j=0}^{k} z^{-(j+1)} \right\} h_n(z) dz$$

$$= (2\pi i)^{-1} \int_{\mathcal{P}} z^{-(k+1)} (1-z)^{-1} h_n(z) dz, \tag{7.21}$$

where $h_n(z) = \sum_{j=0}^{\infty} p_j z^j = \prod_{\ell=1}^{n} (1+z^\ell)/2$, and $\mathcal{P}$ is any closed path surrounding the origin in the complex plane, but not the point $z = 1$. By defining

$$\tilde{w}_n(z) = \log h_n(z) - (k+1)\log z - \log(1-z),$$

the right hand side of (7.21) assumes the form of the integral (3.1), where the "large parameter" $v(= n)$ is already included in $\tilde{w}_n(z)$. At this point Helstrom (1986b) computes numerically the saddlepoint z_0 which satisfies

$$\tilde{w}_n'(z_0) = 0,$$

and the path of steepest descent given by

$$\mathcal{I}\tilde{w}_n(z) \equiv 0,$$

and he integrates along this path by using the trapezoidal rule; cf. Rice (1973). We refer to that paper for the numerical aspects.

Exhibit 7.8 gives the relative error $(= (q_k(\text{SAD}) - q_k(\text{exact}))/q_k(\text{exact}))$ for several sample sizes and probability levels. These results show once more the great accuracy of the saddlepoint approximation even far out in the tail. Similar results are obtained for the

two-sample case; see Helstrom (1986b).

	Probability Level					
		0.00005		0.0005		0.005
Sample Size	N_s	Rel. Error	N_s	Rel. Error	N_s	Rel. Error
20	18	-0.27(-2)	16	-0.82(-3)	14	0.15(-4)
	35	-0.27(-2)	31	-0.81(-3)	27	0.15(-4)
25	20	-0.17(-3)	18	-0.71(-5)	16	0.58(-5)
	40	-0.13(-3)	36	-0.74(-5)	32	0.57(-5)
30	23	0.44(-5)	21	-0.14(-5)	17	0.13(-6)
	45	0.37(-5)	41	-0.14(-5)	31	-0.19(-7)
35	25	-0.55(-6)	18	-0.93(-7)	17	0.16(-6)
	50	-0.63(-6)	36	-0.98(-7)	33	-0.14(-7)
40	21	-0.44(-7)	18	0.22(-8)	17	0.20(-6)
	42	-0.45(-7)	35	-0.74(-8)	33	-0.17(-8)

	Probability Level					
		0.01		0.05		0.2
Sample Size	N_s	Rel. Error	N_s	Rel. Error	N_s	Rel. Error
20	13	0.22(-4)	12	-0.30(-5)	11	0.56(-4)
	26	0.21(-4)	24	-1.00(-5)	22	-0.44(-6)
25	15	-0.83(-6)	14	0.44(-5)	9	0.56(-4)
	30	-0.12(-5)	28	0.57(-6)	18	-0.29(-6)
30	16	0.23(-6)	16	0.51(-5)	8	0.55(-4)
	29	-0.20(-6)	32	-0.18(-7)	15	-0.11(-8)
35	17	0.48(-6)	9	0.53(-5)	7	0.55(-4)
	33	0.12(-7)	17	-0.34(-8)	14	0.54(-9)
40	11	0.55(-6)	8	0.56(-5)	7	0.55(-4)
	19	-0.11(-8)	15	-0.33(-9)	14	0.47(-9)

Exhibit 7.8 (from Helstrom, 1986b)

Relative error $(= (q_k(\text{SAD}) - q_k(\text{exact}))/q_k(\text{exact})$ for the distribution of the one-sample Wilcoxon statistic. N_s is the number of steps in the numerical integration involved. 0.22(−4) means $0.22 \cdot 10^{-4}$.

7.5. APPLICATIONS IN COMMUNICATION THEORY AND RADAR DETECTION

In signal detection problems engineers work with very small false-alarm probabilities. Accurate real time approximations of extreme tail areas are therefore required in these situations; cf. Exhibit 2.3. There is a growing number of papers in the engineering literature where the method of steepest descent and saddlepoint techniques are used to find good approximations to tail areas. In this section we summarize the key ideas with two typical examples.

The first example is taken from Helstrom and Ritcey (1984).

Consider a receiver that integrates a large number n of pulses. The problem of radar detection of a known signal in additive white Gaussian noise can be formulated as a testing problem

$$H_0 : T_n = \frac{1}{2}\sum_{j=1}^{n}(x_j^2 + y_j^2),$$

against

$$H_1 : T_n = \frac{1}{2}\sum_{j=1}^{n}[(x_j + s_j)^2 + (y_j + t_j)^2], \tag{7.22}$$

where $x_1, \cdots, x_n$, $y_1, \cdots, y_n$ are n iid Gaussian random variables with zero mean and unit variance and $s_j^2 + t_j^2 = |d_j|^2$. The $d_j = s_j + it_j$ are the complex amplitudes of the target echoes and the total signal-to-noise ratio is $S = \frac{1}{2}\sum_{j=1}^{n}|d_j|^2$. The d_j's are fixed for a steady target and are random variables when the target fluctuates. In the latter case the phase of the amplitude is uniformly distributed on $[0, 2\pi]$. According to the Neyman-Pearson Lemma the system decides H_0 or H_1 according to whether

$$T_n \underset{H_0}{\overset{H_1}{\gtrless}} t_0,$$

where t_0 is the critical value depending on the false-alarm probability (level of the test). Note that the test based on this test statistic is not robust and T_n can be modified to be resistant to outlying observations. However, our goal here is to investigate the distribution of T_n.

No simple analytic expression is available for the density of the test statistic T_n especially in the case of a fluctuating target. Moreover, the computation of t_0 for very small false-alarm probabilities and the computation of the detection probability (power of the test) requires the integration of the density of T_n far out in the tails. In this case the normal approximation is of little help; cf. Exhibit 2.3.

In many electrical engineering applications, although the density in general is difficult to evaluate, the moment generating function is easy to determine. From Helstrom and Ritcey (1984) the moment generating function for a nonfluctuating target (for a fixed signal-to-noise ratio S) is

$$M_n(\alpha|S) = (1-\alpha)^{-n}\exp\{S\alpha/(1-\alpha)\}$$

and that for a fluctuating target is

$$M_n(\alpha) = (1-\alpha)^{-n} M_n^{(S)}(-\alpha/(1-\alpha)),$$

where $M_n^{(S)}$ is the (known) moment generating function of S. Then, by the same arguments as in section 3.3 or section 4.3, one can write the density of T_n as

$$f_n(t) = \frac{1}{2\pi i} \int_{\tau-i\infty}^{\tau+i\infty} M_n(z) \exp(-zt) dz \tag{7.23}$$

and the false-alarm probability

$$\int_{t_0}^{\infty} f_n(t) dt = \frac{1}{2\pi i} \int_{\tau-i\infty}^{\tau+i\infty} M_n(z) \exp(-zt_0) dz \tag{7.24}$$

cf. also (6.1).
The detection probability can be computed in a similar way.

At this point the integrals in (7.23) and (7.24) could be evaluated by using the saddlepoint techniques discussed in the previous chapters. However, engineers rewrite the integral in (7.24) as $\int_{\tau-i\infty}^{\tau+i\infty} \exp[\tilde{w}_n(z)] dz$, where $\tilde{w}_n(z) = -\log z - zt_0 + K_n(z)$, with $K_n(z) = \log M$, and determine the saddlepoint z_0 and the corresponding path of steepest descent. Now, instead of developing the exponent in a series around the saddlepoint and keeping the leading term, they evalute numerically the integral along the contour of integration defined in the x-y plane by $\Im w_n(x+iy) = 0$. This contour corresponds to the path of steepest descent from the saddlepoint z_0; see section 3.2. The numerical integration is performed using the trapezoidal rule as developed in Rice (1973). Since the contour of integration is a curve on the x-y plane defined implicitly, it is often approximated by a parabola crossing the real axis at the (real) saddlepoint z_0.

Exhibit 7.9 shows a comparison of saddlepoint approximations and numerical integration techniques for some selected cases. The integration's contours considered in this exhibit are the "vertical contour" and the "parabolic contour". The former is obtained by approximating the exact integration's contour by a vertical straight line at the saddlepoint whereas the latter is obtained by using a quadratic approximation of the integration path at the saddlepoint. Exhibit 7.9 shows the improvement of the accuracy by using the numerical integration and the faster convergence of the parabolic contour method compared to the vertical one.

$\bar{S}/n$ in dB	Saddlepoint approximation	Numerical contour integration: Vertical contour	# steps	Numerical contour integration: Parabolic contour	# steps
		$n = 10, \kappa = 20, t_0 = 32.717$			
$-\infty$	1.0008(−6)*	9.971482(−7)*	16	9.951845(−7)*	5
		9.951150(−7)*	32	9.951149(−7)*	10
5.0	2.1588(-1)	2.190677(-1)	12	2.186425(-1)	6
		2.184155(-1)	24	2.184151(-1)	11
		$n = 100, \kappa = 200, t_0 = 154.9$			
$-\infty$	1.0064(−6)*	1.007402(−6)*	6	1.007217(−6)*	5
		1.007099(−6)*	11	1.007099(−6)*	10
-2.3	3.8960(-1)	4.100784(-1)	7	4.097784(-1)	6
		4.086649(-1)	13	4.086641(-1)	12
0.0	4.9742(-3)	4.993271(-3)	6	4.991731(-3)	5
		4.990710(-3)	11	4.990710(-3)	10
		$n = 500, \kappa = 1000, t_0 = 613.576$			
$-\infty$	9.9900(−7)*	1.000187(−6)*	5	1.000160(−6)*	5
		1.000041(−6)*	10	1.000041(−6)*	10
-7.0	2.8300(−1)*	2.989386(−1)*	6	2.990300(−1)*	6
		2.982982(−1)*	12	2.982983(−1)*	12
-5.0	5.8950(-2)	5.985495(-2)	6	5.984295(-2)	6
		5.980515(-2)	11	5.980515(-2)	11

Exhibit 7.9
(from Helstrom and Ritcey (1984))

False dismissal probability (type II error probabilty) for radars with fluctuating target (Swerling Case IV Target) computed by means of saddlepoint techniques and numerical integration on two approximation of the contour (vertical contour and parabolic contour).

$\bar{S}$ is the average signal-to-noise ratio and $\bar{S}/n$ is given in dB $(= 10\log_{10}(\cdot))$. $\kappa = 2n$ is the parameter of the distribution of S for Swerling Case IV Targets. 5.8950(−2) means $5.8950 \cdot 10^{-2}$. An asterisk indicates that 1 - false dismissal probability (= detection probability = power of the test) is reported.

The second example is taken from Helstrom (1986a).

Consider a binary symmetric channel with intersymbol interference. The output of the receiver at time t has the form

$$x(t) = s(t) + \epsilon(t), \tag{7.25}$$

where the signal $s(t)$ is given by

$$s(t) = \sum_{j=-\infty}^{+\infty} b_j a(t - jT)$$

and $\epsilon(t)$ is the noise with mean zero and a symmetric density such as the normal. The b_j's are $+1$ or -1 with equal probability and express the transmitted message. They are assumed uncorrelated. $a(t)$ is the pulse transmitted and T is the interval between pulses. By sampling at a time t we obtain

$$x = \sum_{j=-\infty}^{+\infty} b_j a_j + \epsilon, \qquad a_j = a(t - jT).$$

The receiver then decides whether a particular one of the b_j, say b_0, is $+1$ or -1 by choosing $+1$ if $x > 0$ and -1 if $x < 0$. The error probability is given by

$$P_{err} = P\big[x = \eta + a_0 + \epsilon < 0 | b_0 = +1\big],$$

where

$$\eta = \sum_{\substack{j=-\infty \\ j\neq 0}}^{+\infty} b_j a_j.$$

The exact calculation of P_{err} is very difficult because all possible combinations of the $b_j = \pm 1$ must be taken into account, $1 \leq |j| \leq \infty$. Even with only a finite number n of pulses before and after a_0, the computation of the error probability involves 2^{2n} combinations, where n is large in practical applications.

As in the previous example the moment generating function of x can be determined easily. When a finite number $2n$ of pulses is considered to interfere with b_0, the moment generating function is given by

$$M_n(\alpha) = \gamma_n(\alpha) M_n^{(\epsilon)}(\alpha) \exp(a_0 \alpha), \tag{7.26}$$

where $M_n^{(\epsilon)}$ is the (known) moment generating function of the noise ϵ and

$$\gamma_n(\alpha) = E\big[e^{\eta \cdot \alpha}\big] = E\Big[\exp(\alpha \sum_{\substack{j=-n \\ j\neq 0}}^{n} b_j a_j)\Big]$$

$$= \prod_{\substack{j=-n \\ j\neq 0}}^{n} \cosh(a_j \alpha). \tag{7.27}$$

Then, by the same arguments as in (7.23) and (7.24), we can rewrite the density of x as

$$f_n(x) = \frac{1}{2\pi i} \int_{\tau - i\infty}^{\tau + i\infty} M_n(z) e^{-zx} dz$$

and the error probability

$$P_{err} = \int_{-\infty}^{0} f_n(x)dx = \frac{1}{2\pi i}\int_{r-i\infty}^{r+i\infty} M_n(z)\left\{\int_{-\infty}^{0} e^{-zx}dx\right\}dz$$

$$= -\frac{1}{2\pi i}\int_{r-i\infty}^{r+i\infty} z^{-1}M_n(z)dz. \tag{7.28}$$

The integral in (7.28) can now be evaluated by using saddlepoint techniques and numerical contour integration; see Helstrom (1986a).

Further engineering applications of saddlepoint echniques can be found in Helstrom (1978, 1979, 1985), Helstrom and Rice (1984), and Ritcey (1985).

7.6. OTHER APPLICATIONS OF THE SADDLEPOINT METHOD

After Daniels (1954) pioneering paper, saddlepoint techniques have been applied successfully to several types of problems. Several applications are already given in the previous chapters. In this section we summarize some further applications, the goal being to give an idea of the diversity of situations where these techniques can be used.

Keilson (1963) applied saddlepoint techniques to find an approximation to the density of the sum of N iid random variables, where N is a Poisson process, whereas Blackwell and Hodges (1959) and Petrov (1965) used these ideas to derive approximations to tail probabilities of the sum of iid random variables. A generalization of Keilson's paper can be found in Embrechts et al. (1985). A computation is given in Exhibit 3.12.

Although saddlepoint methods rely on the existence of the moment generating function, Daniels (1960) shows a case where the technique can be carried out when this condition fails.

The approximation to the density of the serial correlation coefficient is considered in Daniels (1956) and Durbin (1980b). Moreover, Daniels (1982) uses saddlepoint approximations in birth processes. These papers show that these techniques can be applied in non iid problems. A related paper is Bolthausen (1986) where Laplace approximations for Markov processes are discussed.

An important application is considered by Robinson (1982). In this paper tail areas for permutation tests are approximated and by inversion approximate confidence intervals are constructed; cf. also the short discussion in Daniels (1955).

The steepest descent method combined with numerical integration is used by Imhof (1961), Lugannani and Rice (1984), and Helstrom and Ritcey (1985) to approximate the density of quadratic forms of normal variables and noncentral F-distributions.

An application of similar ideas in the Bayesian framework is provided by Tierney and Kadane (1986). Basically, Laplacian techniques are used to approximate the integrals in the numerator and the denominator when computing a posterior expectation. Software to perform these computations is available in XLISP-STAT; see Tierney (1990).

Important applications in econometrics can be found in Phillips (1978) and Holly and Phillips (1979). In the latter paper the density of the k-class estimator in the case of the equation with two endogeneous variables is approximated. A related paper is Chesher, Peters, Spady (1989).

Further examples are given in Good (1957, 1961), Keilson (1965), Barndorff-Nielsen and Cox (1979), and Davison and Hinkley (1988), and Johns (1988).

7.7. OUTLOOK

In this monograph we have shown how small sample asymptotic techniques can be successfully applied to many different types of problems. At different places we have pointed to possible open problems and research directions. Clearly, the striking characteristic of these approximations is the great accuracy down to very small sample sizes. But, in spite of the accumulated numerical evidence on many different problems and in spite of the theoretical results on the expansions, the answer to the question as to *why small sample asymptotics does work so well for very small sample sizes* is still not completely satisfactory. More research probably together with the development of new analytic tools to measure this accuracy is needed.

Small sample asymptotics has now reached the point where it can be implemented in computer packages. Both the practitioner and the researcher would benefit greatly from this software development.