Efficiency and deficiency considerations in the symmetry problem

by W. Albers *

Summary Usually, two statistical procedures A and B are compared by means of their asymptotic relative efficiency $e$. If $e = 1$, however, it is more informative to compare A and B by means of the concept of deficiency, which was introduced by Hodges and Lehmann [7]. In the present paper we use this concept for the comparison of linear rank tests and parametric tests for the symmetry problem. In this problem, the hypothesis has to be tested that a sample comes from a distribution that is symmetric about zero. The results provide new and strong evidence for the nice performance of linear rank tests for the symmetry problem. The present paper gives a survey of the results obtained by Albers, Bickel and van Zwet [1] and by Albers [2].

1. Efficiency and deficiency

Before considering the symmetry problem, we shall introduce “efficiency” and “deficiency” in a general context. Let A and B be two statistical procedures for the same problem. Assume that we agree on a criterion by which the performance of A and B is measured. If A and B are tests, this criterion will usually be the power of the tests, if A and B are unbiased estimators, it will usually be the variance of the estimators, etc. For $N = 1, 2, \ldots$ we now define $k_N$ in the following way: if A is based on $N$ observations then $k_N$ is the number of observations which is needed for B to attain the same level of performance as A.

A possible way to compare A and B is to study the behaviour of the ratio $e_N = N/k_N$. In general, it is not possible to find $e_N$ for fixed values of $N$, as the exact values of the performance criteria of A and B are usually not known. In many cases, however, asymptotic results about these criteria are available and these enable us to find $e = \lim_{N \to \infty} e_N$, if it exists. The limit $e$ is called the asymptotic relative efficiency of B with respect to A. Computations of this type are by now almost classical: as early as 1925 Fisher [5] found $e = 2/\pi$ in comparing the median and the mean for the estimation of normal location.

Another way to compare A and B is to consider the behaviour of the difference $k_N - N$. Although this difference seems to be a very natural quantity to examine, historically the ratio $N/k_N$ was preferred by almost all authors in view of its simpler behaviour. The first general investigation of $k_N - N$ was carried out by Hodges and Lehmann [7]. They name $k_N - N$ the deficiency of B with respect to A and denote it as $d_N$. If $\lim_{N \to \infty} d_N$ exists, it is called the asymptotic deficiency of B with respect to A and denoted as $d$.

As concerns the relation between efficiency and deficiency, we distinguish two cases. In the first place, if $e \neq 1$, it does not make much sense to consider the deficiency

---

* Technological University Twente, Enschede.
cy: if $N$ tends to infinity, $d_N$ will tend to infinity at the same rate as $N$. If $e = 1$, however, the situation is entirely different. In this case, $A$ and $B$ perform equally well in first order, and from the fact that $e = 1$ we cannot even deduce which of the two procedures is the better one. Hence, to be able to judge the difference in performance between $A$ and $B$, we have to apply a more refined measure here and it becomes interesting to consider $d_N$. In view of this, we restrict the study of deficiencies to the case where $e = 1$, which occurs in many important statistical problems.

Under the assumption $e = 1$ we evaluate $d_N$ and $d$ in the following way. Denote the performance criteria for $A$ and $B$ as $P_{A,N}$ and $P_{B,N}$, respectively. By definition, $d_N = k_N - N$ may, for each $N$, be found from

$$P_{A,N} = P_{B,k_N}.$$  \hspace{1cm} (1.1)

In order to solve (1.1), $k_N$ has to be treated as a continuous variable. There are several ways to do this in a satisfactory manner. One possibility is the following: let $[y]$ denote the integer part of $y$. Then, for non-integer $k_N$, we select sample size $[k_N]$ or $[k_N] + 1$ with probability $1 - k_N + [k_N]$ and $k_N - [k_N]$, respectively, thus yielding a continuous expected sample size $k_N$. The performance of $B$ is then measured by

$$P_{B,k_N} = (1 - k_N + [k_N])P_{B,[k_N]} + (k_N - [k_N])P_{B,[k_N] + 1}.$$  \hspace{1cm} (1.2)

As we already mentioned, $P_{A,N}$ and $P_{B,N}$ are generally not known exactly and we have to use asymptotic results. To find the asymptotic relative efficiency of $B$ with respect to $A$ it suffices to have an asymptotic result of the following kind

$$P_{A,N} = \frac{c_1}{N^r} + o\left(\frac{1}{N^r}\right), \quad P_{B,N} = \frac{c_2}{N^r} + o\left(\frac{1}{N^r}\right).$$  \hspace{1cm} (1.3)

for certain $c_1$ and $c_2$ not depending on $N$ and for a positive constant $r$. From (1.1) and (1.2) it then follows that $(N/k_N)^{-1} = (c_1/c_2) + o(1)$ and therefore $e = (c_1/c_2)^{1/r}$. In particular, if $c_1 = c_2$, we have $e = 1$.

For the evaluation of deficiencies, the information contained in (1.2) is not sufficient; we need a stronger result, namely

$$P_{A,N} = \frac{c}{N^r} + \frac{a}{N^{r+s}} + o\left(\frac{1}{N^{r+s}}\right),$$  

$$P_{B,N} = \frac{c}{N^r} + \frac{b}{N^{r+s}} + o\left(\frac{1}{N^{r+s}}\right).$$  \hspace{1cm} (1.3)

for certain $c$, $a$ and $b$ not depending on $N$ and certain positive constants $r$ and $s$. The leading term of both expansions in (1.3) is chosen to be equal since we are only interested in deficiencies if $e = 1$. From (1.1) and (1.3) we can solve $d_N$:  

82
which implies that

\[ d_N = \frac{(b-a)}{rc} N^{(1-s)} + o(N^{(1-s)}) \]  

(1.4)

Hence

\[
\begin{cases}
\pm \infty, & 0 < s < 1, \\
\frac{b-a}{rc}, & s = 1, \\
0, & s > 1.
\end{cases}
\]  

(1.5)

From this derivation it appears that the real problem in the evaluation of \( d \) is to establish the asymptotic expansions in (1.3). As soon as those have been obtained, \( d_N \) and \( d \) follow immediately from (1.4) and (1.5). In order to give a simple demonstration of the use of deficiency we quote an example due to HODGES and LEHMANN [7], in which the asymptotic expansions are readily available. Consider a sample \( X_1, \ldots, X_N \) from a distribution \( F \) with expectation \( \xi \) and variance \( \sigma^2 \). Then two frequently used estimators of \( \sigma^2 \) are

\[
M_N = \frac{1}{N} \sum_{i=1}^{N} (X_i - \xi)^2
\]

and

\[
M' = \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \bar{X})^2, \quad \text{where} \quad \bar{X} = \frac{1}{N} \sum_{i=1}^{N} X_i.
\]

Here \( M_N \) is used if \( \xi \) is known and \( M'_N \) is used if this is not the case or if we do not dare to rely on its given value. Both estimators are unbiased and therefore we use the variance as criterion of performance. Define \( \gamma \) by \( \gamma + 1 = \mu_4/\sigma^4 \), where \( \mu_4 \) is the fourth central moment of \( F \).

Then we have

\[
\sigma^2(M_N) = \sigma^4 \frac{\gamma}{N},
\]

\[
\sigma^2(M'_N) = \sigma^4 \frac{\gamma(N-1)+2}{N(N-1)} = \sigma^4 \left( \frac{\gamma}{N} + \frac{2}{N^2} + o\left( \frac{1}{N^3} \right) \right).
\]  

(1.6)

(These results follow by straightforward computation; see CRAMER [3], formula
(27.4.2), for $\sigma^2 (M,\phi)$. If $\gamma > 0$, the results in (1.6) are of the form (1.3) and application of (1.4) and (1.5) shows that here the asymptotic deficiency $d = 2/\gamma$. If $F$ is normal, $\gamma = 2$ and hence $d = 1$: the price of not knowing the mean is asymptotically one additional observation. Note that in the normal case not only $d = 1$, but that also $d_N = 1$, since $M_N$ and $M_{N+1}$ are identically distributed. In practice one frequently encounters distributions having somewhat heavier tails than the normal distribution. For such distributions $\gamma$ will in general be larger than 2, and hence $d$ will even be smaller than 1 here.

The remaining part of the paper is devoted to the application of the deficiency concept to various tests for the symmetry problem. Here the derivation of asymptotic expansions for the performance criteria leads to considerable difficulties. In the next section we formulate the problem.

2. Comparison of tests for the symmetry problem

Suppose that we want to test the effect of a treatment by comparing $N$ pairs of subjects which have been matched such as to eliminate as far as possible any differences not due to the treatment. One member of each pair is chosen at random to receive the treatment, while the other serves as control. Let $(Y_1, Z_1), \ldots, (Y_N, Z_N)$ be the resulting sample. Then the hypothesis of no treatment effect can be reduced by intuition to the hypothesis $H_0$ that the distribution function (d.f.) $G$ of the differences $X_i = Y_i - Z_i$, $i = 1, \ldots, N$ is symmetric about zero, i.e.

$$H_0: G(-x) = 1 - G(x) \text{ for all } x.$$  

The hypothesis $H_0$ is called the hypothesis of symmetry; the problem of testing $H_0$ is the symmetry problem (cf. LEHMANN [8], § 6.7 and § 6.9).

One often considers a simplified version of this problem by assuming that $G$ belongs to a parametric family of the form

$$\left\{ F|F(x) = F_1 \left( \frac{x-\theta}{\sigma} \right), F_1 \text{ known and symmetric about zero} \right\}.$$  

Then $H_0$ reduces to

$$H'_0: \theta = 0.$$  

The most common choice of $F_1$ is of course $F_1 = \Phi$, the standard normal d.f.. Well-known tests for $H_0$ in the normal case are the one-sample $t$-test and the test based on the sample mean $\bar{X} = N^{-1}\sum_{i=1}^N X_i$. These tests have nice optimality properties if the assumption of normality is correct. If this is not the case, however, the tests are no longer valid and may lead to entirely wrong results.

In view of this, one might prefer tests which are valid without assumptions about
the type of the distribution and which can be used to test $H_0$, rather than merely $H_0^*$ for one particular choice of $F_1$. Such tests are called non-parametric or distribution-free: the distribution of their test statistics has to be the same for all symmetric d.f.'s $G$.

A well-known class of tests satisfying this requirement is the class of one-sample linear rank tests. These are introduced as follows: take the absolute values $|X_1|, \ldots, |X_N|$ of $X_1, \ldots, X_N$ and rearrange these absolute values such that they increase. Let us denote this ordered sequence by $0 < Z_1 < \ldots < Z_N$. Furthermore, we introduce $V_1, \ldots, V_N$ defined by

$$V_j = \begin{cases} 1 & \text{if the } X_i \text{ corresponding to } Z_j \text{ is positive}, \\ 0 & \text{otherwise}. \end{cases} \quad (2.1)$$

Finally, let $a = (a_1, \ldots, a_N)$ be a vector of scores and define

$$T = \sum_{j=1}^{N} a_j V_j. \quad (2.2)$$

Every test for the symmetry problem that is based on a statistic of the form (2.2), is called a linear rank test. The simplest example of such a test is obtained by taking $a_j = 1, j = 1, \ldots, N$. Then $T$ is the number of positive observations in the sample and the corresponding test is the sign test. Another well-known example occurs for $a_j = j, j = 1, \ldots, N$: this gives Wilcoxon's signed rank test. As a third and final example we mention the one-sample normal scores test, which is obtained by taking $a_j = E|Z_j|, j = 1, \ldots, N$, where $0 < Z_1 < \ldots < Z_N$ are the ordered absolute values of a sample of size $N$ from the standard normal distribution.

The fact that the linear rank tests introduced above are valid for all symmetric d.f.'s $G$ means of course an advantage over tests like the $t$-test and the $X$-test. On the other hand, it seems plausible that, as a price for their wider validity, the former will be less powerful than the latter. Clearly, the choice between the two types of tests will heavily depend on the height of this price. In view of this it seems interesting to know efficiencies and deficiencies of linear rank tests with respect to tests like the $t$-test and the $X$-test. Before we can evaluate these, we have to give a more precise formulation of the circumstances under which we want to compare the two types of tests.

Consider again a parametric family

$$\mathcal{F}_1 = \left\{ F \mid F(x) = F_1 \left( \frac{x-\theta}{\sigma} \right), \; F_1 \text{ known and symmetric about zero} \right\}. \quad (2.3)$$

Let $\psi_S$ be a test for $H_0^*: \theta = 0$, based on a statistic $S$. The most general alternative is of course $(H_0^*)^C$, but for simplicity we restrict attention to the one-sided alternative hypothesis

$$H_1^*: \theta > 0.$$
Let $\psi_T$ be a linear rank test for $H_0$, based on a statistic $T$ of the form (2.2). Then we want to compare the performance of $\psi_T$ and $\psi_S$ under $H_1$. As our performance criteria we choose the power functions of $\psi_S$ and $\psi_T$, which we denote as $\pi_S$ and $\pi_T$, respectively.

In order to find the efficiency of $\psi_T$ with respect to $\psi_S$ we have to determine the behaviour of $\pi_S$ and $\pi_T$ as the sample size $N \to \infty$. For a fixed test size $\alpha$ and a fixed alternative $\theta$, the power of every reasonable test will tend to 1 as $N \to \infty$. Typically, $\psi_S$ and $\psi_T$ will be such reasonable tests and therefore we will find $\lim_{N \to \infty} \pi_S = \lim_{N \to \infty} \pi_T = 1$.

This result is not sufficiently informative for the evaluation of efficiencies. Such an evaluation would require knowledge of the rate of convergence of $\pi_S$ and $\pi_T$ towards 1. This, however, is a complicated matter and for linear rank tests very little is known about it. Therefore, generally the following approach is used: the test size $\alpha$ remains fixed but instead of a fixed alternative $\theta$ we consider so-called local or contiguous alternatives. This means that we look at sequences of alternatives $\{\theta_N\}$ for which $\theta_N \to 0$ as $N \to \infty$ at such rate that the power tends to a limit which lies strictly between $\alpha$ and 1.

It can be shown that the class of these sequences is usually the class of sequences $\{\theta_N\}$ for which $\lim_{N \to \infty} N^\alpha \theta_N = b$, for some constant $b$ with $0 < b < \infty$. For such sequences, methods are available to find $\lim_{N \to \infty} \pi_T(N, \alpha, \theta_N)$ and $\lim_{N \to \infty} \pi_S(N, \alpha, \theta_N)$. By comparing these limits we can evaluate the asymptotic relative efficiency $e_{T,S}$ of $\psi_T$ with respect to $\psi_S$ for fixed test sizes $\alpha$ and local alternatives

$$H_1^{1N}: \theta_N > 0, \lim_{N \to \infty} N^\alpha \theta_N = b, 0 < b < \infty. \tag{2.4}$$

Clearly $e_{T,S}$ has a local character; it is sometimes called Pitman-efficiency to avoid confusion with other efficiency measures.

Next we specialize to the case where $T$ and $S$ are optimal in some sense, which is of course of particular interest. It can be shown that, for every parametric family $\mathcal{F}_1$ from (2.3), there exists a test $\psi_S$, which is asymptotically most powerful among all tests for $H_0: \theta = 0$ against $H_1^{1N}$. Likewise there exists, for each $\mathcal{F}_1$, a linear rank test $\psi_T$, which is asymptotically most powerful among all linear rank tests for the hypothesis of symmetry $H_1^{1N}$ against $H_1^{1N}$. For example suppose that

$$\mathcal{F}_1 = \left\{ F|F(x) = \Phi \left( \frac{x - \theta}{\sigma} \right) \right\},$$

where $\Phi$ is the standard normal d.f. If the scale parameter $\sigma$ is known, $\psi_S^*$ is the $X$-test; if $\sigma$ is unknown, $\psi_S^*$ is the $t$-test. In both cases $\psi_T^*$ is the normal scores test. Obviously, $\psi_T^*$ cannot be better than $\psi_S^*$, as the linear rank tests for $(H_0, H_1^{1N})$ form a sub-class of the class of all tests for $(H_0, H_1^{1N})$. In particular $e_{T^*, S^*}$ will certainly satisfy $e_{T^*, S^*} \leq 1$. It can be shown, however, that the upper bound is always attained:
$e_{\tau^*, S^*} = 1$, (cf. HÁJEK and SIĐÁK [6]). This important result means that under $H_{1\mu}$ the use of a linear rank test does not lead to a loss of efficiency, at least asymptotically to first order.

At this point it appears that the present problem is of the type we discussed in general in section 1: we have two procedures $\psi_{\tau^*}$ and $\psi_{S^*}$ for the same problem, and $e_{\tau^*, S^*} = 1$. Hence we would like to know the asymptotic deficiency $d_{\tau^*, S^*}$ of $\psi_{\tau^*}$ with respect to $\psi_{S^*}$. From section 1 it is known that the evaluation of $d_{\tau^*, S^*}$ boils down to the derivation of asymptotic expansions for the power functions $\pi_{\tau^*}$ and $\pi_{S^*}$. This problem is considered in the next section.

3. Asymptotic expansions for the power of linear rank tests

Here we shall give an idea of the way in which asymptotic expansions for the power of linear rank tests for the symmetry problem can be obtained. For an exact and detailed derivation of such expansions the interested reader is referred to ALBERS, BICKEL and VAN ZWET [1] and to ALBERS [2].

Let $\psi_{\tau^*}$ be a linear rank test for $H_0$, based on a statistic $T$ of the form (2.2). To find an expansion for the power of $\psi_{\tau^*}$, we need an expansion for the distribution function (d.f.) of $T$, both under hypothesis and alternative. For convenience, we standardize $T$ and look for an expansion $\bar{R}(x)$ for the d.f. $R(x)$, defined as

$$R(x) = P\left( \frac{T - ET}{\sigma(T)} \leq x \right) \quad (3.1)$$

Our starting point in establishing $\bar{R}(x)$ will be the so-called Edgeworth expansions. For the d.f. $H(x)$ of any random variable (r.v.) $Y$ with mean 0 and variance 1 we can give a formal Edgeworth expansion $\bar{H}(x)$ in powers of $N^{-\frac{1}{2}}$. For our purposes it will suffice to consider the terms up to $O(N^{-\frac{1}{2}})$. Then $\bar{H}(x)$ looks like

$$\bar{H}(x) = \Phi(x) - \varphi(x) \left[ \frac{N^{-\frac{1}{2}}K_3}{6} (x^2 - 1) + \frac{N^{-\frac{1}{2}}K_4}{24} (x^3 - 3x) + \frac{N^{-\frac{1}{2}}K_3}{72} (x^5 - 10x^3 + 15x) \right], \quad (3.2)$$

where $\kappa_3 = N^\frac{1}{2} EY^3$, $\kappa_4 = N(EY^4 - 3)$ and $\Phi = \Phi'$ is the standard normal density. The expansion $\bar{H}(x)$ in (3.2) already has a nice explicit form: the real problem does not lie in writing down $\bar{H}(x)$ but in justifying it, i.e. in showing that $\kappa_3$ and $\kappa_4$ are bounded functions of $N$ and that

$$\sup_x |H(x) - \bar{H}(x)| = o(N^{-\frac{1}{2}}) \quad (3.3)$$

The standard conditions on the r.v. $Y$ under which this holds, are given by FELLER [4]
and CRAMER [3]. Obviously, these conditions include that $Y$ is non-degenerate and that it possesses a sufficiently high moment. Moreover, it is required that $Y$ has the form $\sum_{j=1}^{n} Y_j$ where the r.v.'s $Y_1, \ldots, Y_n$ are independent identically distributed and have no lattice distributions. As concerns the last condition, we say that a r.v. $Y_i$ has a lattice distribution if there exist real numbers $a$ and $h$ with $h > 0$ such that each value of $Y_i$ can be represented as $y_i = a + vh$ for some integer $v$.

Except for the condition that the $Y_j$ are non-lattice, the conditions above are of the same kind as those of the central limit theorem, which asserts that $H(x) = \Phi(x) + o(1)$. Note that this result is implied by (3.2) and (3.3): the Edgeworth expansions provide refinements of the central limit theorem.

Next, we want to apply these general results about Edgeworth expansions to the d.f. $R(x)$ in (3.1). It appears, however, that $T$ does not satisfy very well the standard conditions above: $T$ can be written as a sum $\sum_{j=1}^{n} a_j V_j$, but the summands $a_j V_j$ are not independent, not identically distributed and they do have lattice distributions. In the following we shall give an idea how these difficulties can be overcome.

The first problem we consider is the fact that the $a_j V_j$ are not identically distributed. This complication is not essential and can be dealt with relatively easy. Results like (3.2) and (3.3) can be derived for this case using the same type of arguments as in the identically distributed case. Only the notation becomes slightly more involved and the conditions on the moments must hold uniformly for all summands.

A more delicate complication is caused by the lattice character of the $a_j V_j$. To understand why this lattice character leads to trouble in establishing Edgeworth expansions, we first consider the special case where $a_j = 1$ for $j = 1, \ldots, N$. Then $\psi_T$ is the sign test and $T$ is the number of positive observations in the sample. Clearly, $T$ has a binomial distribution with parameters $N$ and $p = P(X_i > 0)$. Now let $c$ and $\varepsilon$ be positive constants and let $ET$ and $\sigma^2(T)$ denote the expectation and variance of $T$. Then it follows from the normal approximation to the binomial distribution that $P(|T - ET| \leq \varepsilon \sigma(T)) \rightarrow 2 \Phi(c) - 1 > 0$, as $N \rightarrow \infty$, for $\varepsilon \leq p \leq 1 - \varepsilon$. Hence the binomial distribution places asymptotically a strictly positive mass on an interval of length $2 \varepsilon \sigma(T)$. As the binomial distribution is integer-valued, this mass has to be divided over at most $2 \varepsilon \sigma(T) + 1$ points. But this implies that there is at least one point where the binomial d.f. has a jump which is at least of order $\sigma^{-1}(T)$. From $\sigma^2(T) = Np(1-p)$ it then follows that this jump is at least of order $N^{-\frac{1}{2}}$. However, according to (3.2), Edgeworth expansions are continuous functions. As continuous functions can obviously not approximate functions with jumps of order $N^{-\frac{1}{2}}$ any sharper than to order $N^{-\frac{1}{2}}$, it follows that in the present case (3.3) can never hold. Hence, as far as the sign test is concerned, this approach does not work. Fortunately, the relative simplicity of the distribution of $T$ in this case allows us to establish an expansion for the power of the sign test by other methods (c.f. ALBERS [2]).

If the scores $a_j$ are not all equal, the situation may be different. For then $T$ will attain more values than in the case of the sign test. Hence the probability mass of its distribution can be divided over more points, which may lead to jumps in the d.f. of $T$ that are of sufficiently small order. To illustrate this, we consider Wilcoxon's signed
rank test, where \( a_j = j, j = 1, \ldots, N \). In this case too the distribution of \( T \) is asymptotically normal and moreover \( T \) is also integer-valued. Hence, by the same argument as in the case of the sign test, it follows that the d.f. of \( T \) must have at least one jump of order at least \( \sigma^{-1}(T) \). But here \( \sigma^2(T) \) is of order \( \sum_{j=1}^N j^2 \), i.e. of order \( N^3 \), which implies that at least one jump of order at least \( N^{-\frac{1}{4}} \) occurs. Under these circumstances, an Edgeworth expansion satisfying (3.3), might exist. (In fact it does exist, as is proved by Albers, Bickel and Van Zwet [1]).

The examples above suggest the following conclusions: it is not necessary to require that each summand \( a_j V_j \) itself is non-lattice, as is prescribed by the standard conditions. We only need that the lattice on which the standardized sum \( (T-ET)/\sigma(T) \) is concentrated is sufficiently fine. This can be achieved by imposing a suitable condition on the \( a_j \) which prevents these from getting to close to each other. It can be shown that these conclusions are correct and that the following condition on the \( a_j \) is sufficient: suppose that there exists a constant \( \delta > 0 \) such that \( \lambda \{ x_j | x_j - a_j < \zeta \} \geq \delta N \zeta \) for some \( \zeta \geq N^{-\frac{1}{4}} \log N \), where \( \lambda \) denotes Lebesque measure. This weak condition is satisfied e.g. for the scores of Wilcoxon's signed rank test and for those of the one-sample normal scores test. For the scores of the sign test it is obviously not satisfied.

The last problem we have to deal with, is the fact that the \( a_j V_j \) are not independent. To reveal the nature of this dependence we shall determine the joint distribution of \( V_1, \ldots, V_N \). First we recall briefly the relevant notation from section 2. Let \( X_1, \ldots, X_N \) be \( N \) independent, identically distributed r.v.'s from a d.f. \( G \). Let \( 0 < Z_1 < \ldots < Z_N \) denote the ordered sequence of the absolute values \( |X_1|, \ldots, |X_N| \). Moreover, define \( V_1, \ldots, V_N \) as follows: \( V_j = 1 \) the \( X_i \) corresponding to \( Z_j \) is positive, and \( V_j = 0 \) otherwise, for \( j = 1, \ldots, N \). Finally, denote the corresponding vectors by \( X, V, Z \) and \( V \).

Now one easily verifies that, conditional on \( Z = z \), the r.v.'s \( V_1, \ldots, V_N \) are independent with

\[
P(V_j = 1 | Z = z) = \frac{g(z_j)}{g(z_j) + g(-z_j)}, \quad j = 1, \ldots, N, \quad g = G'
\]

This implies that \( P(V = (1, \ldots, 1)) = E[\prod_{j=1}^N \{g(Z_j)/(g(Z_j) + g(-Z_j))\}] \) and \( \prod_{j=1}^N P(V_j = 1) = E[\prod_{j=1}^N g(Z_j)/(g(Z_j) + g(-Z_j))] \). As \( Z_1 < \ldots < Z_N \) are clearly dependent, it follows that in general \( P(V = 1, \ldots, 1) \neq \prod_{j=1}^N P(V_j = 1) \). Hence \( V_1, \ldots, V_N \) are not independent and we cannot give an Edgeworth expansion directly for the d.f. \( R(x) \) in (3.1). However, as \( V_1, \ldots, V_N \) are independent conditional on \( Z = z \), we can give an Edgeworth expansion \( R(x|z) \) for

\[
R(x|z) = P\left( \frac{T - ET}{\sigma(T)} \leq x|z \right).
\]

By taking the expectation of this conditional expansion with respect to \( Z \) we then find for \( R(x) \) the expansion

\[
R(x) = ER(x|Z).
\]
The result in (3.4) is the formal solution to our problem: after removing several obstacles, we have succeeded in applying the Edgeworth expansions and this has led us to the desired expansion for the d.f. of $T$. It remains, however, to formulate conditions under which this formal result can be translated into a more explicit one. This is a very technical matter which we will leave out of account almost completely here. We shall merely indicate the kind of restrictions under which such a simplification can be achieved. In the first place, if we do not consider general d.f. $G$, but only those which occur under the hypothesis of symmetry or under contiguous alternatives, we can evaluate the expectation in (3.4). For simplicity, we restrict attention to the particular kind of contiguous alternatives given by (2.3) and (2.4), i.e. we consider contiguous location alternatives

$$G(x) = F_1 \left( \frac{x - \theta_N}{\sigma} \right), \quad F_1 \text{ known, } \quad F_1(-x) = 1 - F_1(x) \forall x, \quad \theta_N > 0,$$

$$\lim_{N \to \infty} N^{\frac{2}{3}} \theta_N = b, \quad 0 < b < \infty.$$ (3.5)

Now $R(x|z)$ is a function of $\theta_N$ and $E R(x|z)$ can be evaluated by Taylor expansion with respect to $\theta_N$.

After this, a further simplification of $R(x)$ can be achieved by imposing certain smoothness conditions on the scores $a_j$. Without going into the nature of these conditions, we mention that these are satisfied for example for the scores of Wilcoxon's signed rank test and for those of the one-sample normal scores test.

The resulting simple expansions for the d.f. of $T$ under $H_0$ and under the alternatives (3.5) immediately lead to a simple expansion for the power $\pi_T, F_1(\theta_N)$ of $\psi_T$ under these alternatives. We conclude this section with some examples of such expansions for particular choices of $\psi_T$ and $F_1$. First let $F_1 = \Phi$, the standard normal (NOR) distribution and let $\psi_T$ be the one-sample normal scores (NS) test, which is the asymptotically most powerful linear rank test against normal alternatives of the type (3.5). Then we find

$$\pi_{NS,NOR}(\theta_N) = 1 - \Phi(u_a - \eta_1) - \frac{\eta_1 \varphi(u_a - \eta_1)}{4N} \left\{ u_a^2 - 1 + \log \log N + \log 2 + 0.11664 \ldots \right\} + o(N^{-1}),$$ (3.6)

where $u_a = \Phi^{-1}(1 - \alpha)$ and $\eta_1 = N^{\frac{2}{3}} \theta_N$.

As a second example we consider $F_1(x) = 1/(1 + e^{-x})$, the logistic (L) distribution. For $\psi_T$ we choose Wilcoxon's ($W$) signed rank test, which is the asymptotically most powerful linear rank test against logistic alternatives of the type (3.5). We find

$$\pi_{W,L}(\theta_N) = 1 - \Phi(u_a - \eta_2) - \frac{\eta_2 \varphi(u_a - \eta_2)}{20N} \left\{ 3u_a^2 + u_a \eta_2 + 2 + \eta_2^2 \right\} + o(N^{-1}),$$ (3.7)

where $\eta_2 = (N/3)^{\frac{2}{3}} \theta_N$. Expansions like (3.6) and (3.7), but with slightly more involved coefficients, are also available for $\pi_{NS,L}(\theta_N)$ and $\pi_{W,NOR}(\theta_N)$ (cf. [1] or [2]).
4. Deficiency results for the symmetry problem

In the previous section we established expansions for the power of linear rank tests. Similar expansions can be derived for various other tests for the symmetry problem (c.f. ALBERS [2]). Combining these results we can evaluate deficiencies of linear rank tests with respect to the other tests.

We again consider as our first example the normal case $F_1 = \Phi$. Under these alternatives we want to compare the normal scores test to the test based on the sample mean $\overline{X}$, which is the most powerful test, and to the $t$-test, which is the most powerful scale invariant test. The power $\pi_{\overline{X}, \text{NOR}}(\theta_n)$ of the $\overline{X}$-test is known exactly. It equals

$$\pi_{\overline{X}, \text{NOR}}(\theta_n) = 1 - \Phi(u_{\pi} - \eta_1), \tag{4.1}$$

where $\eta_1 = N^4 \theta_n$. From (3.6) and (4.1) the deficiency $d_n(\text{NS}, \overline{X})$ of the normal scores test with respect to the $\overline{X}$-test is readily found.

As $\{N + d_n(\text{NS}, \overline{X})\}^{1/4} = \eta_1 \{1 + d_n(\text{NS}, \overline{X})/(2N) + o(N^{-2})\}$, the power of the normal scores test based on $N + d_n$ observations, satisfies

$$\pi_{\text{NS}, \text{NOR}}(\theta_n) = 1 - \Phi(u_{\pi} - \eta_1) \frac{\eta_1 \phi(u_{\pi} - \eta_1)}{4N} \left\{ -2d_n(\text{NS}, \overline{X}) + u_{\pi}^2 - 1 + \log \log N + \log 2 + 0.1166 \ldots \right\} + o(N^{-1}).$$

Hence, it follows from the definition of deficiency that

$$d_n(\text{NS}, \overline{X}) = \frac{1}{2}(u_{\pi}^2 - 1) + \frac{1}{2} \log \log N + \frac{1}{2} \log 2 + 0.0583 \ldots + o(1). \tag{4.2}$$

To find the deficiency $d_n(\text{NS}, t)$ of the normal scores test with respect to the $t$-test, we remark that HODGES and LEHMANN [7] have shown that $d_n(t, \overline{X}) = \frac{1}{2}u_{\pi}^2 + o(1)$. As deficiencies are transitive this implies that

$$d_n(\text{NS}, t) = \frac{1}{2} \log \log N + \frac{1}{2} \log 2 - \frac{1}{2} + 0.0583 \ldots + o(1). \tag{4.3}$$

Unfortunately, the asymptotic deficiencies $d(\text{NS}, \overline{X})$ and $d(\text{NS}, t)$ both appear to be infinite. On the other hand, $d_n(\text{NS}, \overline{X})$ and $d_n(\text{NS}, t)$ increase vary slowly as their order is $\log \log N$. To illustrate this, we give some values of the approximation for $d_n(\text{NS}, t)$ in (4.3)

<table>
<thead>
<tr>
<th>$N$</th>
<th>5</th>
<th>10</th>
<th>20</th>
<th>50</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_n(\text{NS}, t)$</td>
<td>0.143</td>
<td>0.322</td>
<td>0.453</td>
<td>0.587</td>
<td>0.668</td>
</tr>
</tbody>
</table>

According to this approximation, even for a sample of size 100, one additional observation suffices to give the normal scores test at least the same power as the $t$-test under normal alternatives. Note that such a result is typically much stronger than the
mere assertion that the ratio of sample sizes needed to attain the same power, tends to 1 as \( N \to \infty \). The present result shows that in the normal case the price for using a distribution-free test is surprisingly low.

As a second example we briefly consider the logistic case, where \( F_1(x) = 1/(1 + e^{-x}) \). Here Wilcoxon's signed rank test is the asymptotically most powerful rank test. For any particular \( \theta_0 \), let \( \psi_{S(\theta_0)} \) be the most powerful test for \( H'_0: \theta = 0 \) against the simple alternative \( \theta = \theta_0 \) and let \( d_N(W, S(\theta_N)) \) denote the deficiency of Wilcoxon's test with respect to \( \psi_{S(\theta_N)} \). It can be shown that \( d_N(W, S(\theta_N)) \) has a finite limit \( d(W, S(\theta_N)) \), which equals

\[
d(W, S(\theta_N)) = \frac{1}{2\theta} (2u + \eta_2)^2 + \frac{3}{1\theta},
\]

where \( \eta_2 = (N/3)\theta_N \). Hence, against logistic alternatives of the type (3.5), a finite number of additional observations suffices to compensate for the amount by which the power of Wilcoxon's test falls short of the maximum power attainable.

Finally, we pay some attention to the question whether the asymptotic results in (3.6), (3.7) and in (4.2)–(4.4) are of value for small to moderate sample sizes. Using exact power results from literature it can be shown that the power expansions (3.6) and (3.7) already yield excellent approximations for samples of size 5–20. The situation becomes entirely different, however, if we have long-tailed distributions under the alternative. For example, in the case of Wilcoxon's signed rank test against Cauchy alternatives the expansion, obtained by the methods of the previous section, leads to very bad results for the same range of sample sizes as above. As concerns the deficiency approximations in (4.2)–(4.4), there appears to exist a satisfactory agreement between the values these approximations yield and those which are derived from the exact power results from literature, again for sample sizes 5–20. More information on finite sample results, including some tables, is given by ALBERS [2].

Acknowledgment

The author is grateful to Professor W. Schafsma for a number of useful comments and suggestions.

References