The Annals of Statistics
1985, Vol. 13, No. 2, 606–625

ROBUST TWO-SAMPLE PERMUTATION TESTS[1]

BY DIANE LAMBERT

Carnegie-Mellon University

A new two-sample randomization test is proposed for testing that the joint distribution of two samples is invariant under permutations. The p-value of the test has a finite sample minimaxity property over neighborhoods of completely specified alternative distributions. Asymptotically, the test has minimax Bahadur slope against the neighborhoods, which remain fixed as the sample sizes increase. The proposed test also offers the best compromise between robustness against departures from a model alternative and optimality at the model alternative in the sense that no other test with the same gross-error-sensitivity has larger slope at the model. Some modifications of the test are proposed for testing the nonparametric null hypothesis against neighborhoods of models that have a shared nuisance location-scale parameter. These nuisance-parameter-free versions of the test are justified for large samples from exponential families, and an example of their use is given.

1. Introduction. Recently there has been renewed interest in permutation tests, also known as randomization or rerandomization tests, as a practical tool for data analysis. Tukey, Brillinger and Jones (1978), for example, encourage their adoption for analyzing weather modification data. Green (1977) has designed an algorithm for computing p-values of permutation tests. Edgington (1980) presents detailed methodology and computer algorithms to enable wider adoption of permutation tests. Gabriel and Hall (1983) give methods for basing confidence intervals on permutation tests. Permutation methods have also received attention under the guise of bootstraps that sample without replacement rather than with replacement (Efron, 1982).

Permutation tests are attractive because the distribution of the observations under the null hypothesis need not be specified to calculate the p-value, and yet the permutation test may be optimal. A two-sample permutation test based on the likelihood ratio of the alternative distributions for two independent random samples is as powerful as any nonparametric test (Lehmann and Stein, 1949). It is also asymptotically as powerful as the best parametric test, having relative efficiency one (Hoeffding, 1952) and deficiency zero in many cases (Bickel and van Zwet, 1978), even if the support of the permutation distribution is only sampled rather than completely enumerated (Vadiveloo, 1983). Moreover, under weak regularity conditions, no other nonparametric p-value is asymptotically

Received October 1982; revised December 1984.

[1] Partial support for the research was provided by NSF grant ATM-7905536 to the University of Rochester and by NSF grant MCS-8301692 to Carnegie-Mellon University.

AMS 1970 *subject classifications.* Primary 62G35; secondary 62E20, 62G10

Key words and phrases. Randomization test, core distribution, minimax p-value, minimax power, minimax slope, influence function, gross-error-sensitivity, data-dependent censoring.

smaller than the likelihood permutation p-value under the alternative (Bahadur et al., 1982).

This optimality at a specific alternative has motivated many authors to recommend that the permutation test be based on the likelihood ratio for an idealized alternative (Box and Andersen, 1955; Kempthorne, 1952; Oden and Wedel, 1975). The recommendation is suspect, however, in light of recent evidence that such a permutation test need not be nearly optimal at distributions near to but not identical to the idealized alternative distributions. The extensive simulations of Higgins and Keller (1979) and Keller (1979), for example, show that the power of the two-sample permutation t-test for samples of sizes 10, 10 deteriorates under contaminated normal, Laplace and Cauchy alternatives. Larger samples do not necessarily increase the robustness of a permutation test. From an asymptotic perspective, the permutation test based on the likelihood ratio lacks qualitative robustness (Lambert, 1982) and has unbounded, weak*-discontinuous influence functions (Lambert, 1981).

Use of the permutation distribution of the test statistic rather than a parametric distribution is not responsible for the lack of test robustness. On the contrary, the influence functions of a two-sample permutation test are always bounded above, even if the influence functions of the corresponding parametric test are unbounded from above and below (Lambert, 1981). Rather than dismiss permutation tests, we should strengthen their tendency toward robustness by choosing a robust test statistic. The major result of this paper is that there is a robust permutation test that is both practical and optimal.

The robust permutation test (RPT) statistic is defined as a sum of censored log-likelihood ratios in Section 2. Because of the censoring, the RPT is insensitive to outliers. Evidence that the RPT resists other departures from the idealized alternative distributions is presented in Sections 4 and 5. Each robustness property is stated for the p-value of the RPT since a p-value rather than an accept-reject decision often summarizes test results.

The RPT is inappropriate for testing for a shift in the presence of shared location and scale nuisance parameters because its censoring points depend on all the parameters of the idealized alternative distributions. In Section 3, RPT's with data-dependent censoring are proposed to accomodate location and scale nuisance parameters in two-sample shift alternatives. For each test, censoring points are first determined from the data, and then a RPT is carried out as if the censoring points had not been determined from the data. Several kinds of data-dependent censoring are discussed. One surprising consequence of censoring at the order statistics of the combined samples (i.e., Winsorizing the combined samples) is that data from the idealized alternative distributions tend to be more heavily censored than data from the worst case alternative distributions.

One new test proposed in Section 3 for an alternative of the type that the df of Y is shifted to the right of the df of X is as follows. Replace each observed X_i and Y_j below $k_1 \equiv X_{[m\beta_1]}$ by k_1 and replace each X_i and Y_j above $k_2 \equiv Y_{[n\beta_2]}$ by k_2, and then carry out the usual permutation test based on averages. The censoring fractions β_1, β_2 may be related to alternatives defined as neighborhoods of idealized or model alternative distributions, and they are independent of location and scale nuisance parameters shared by X and Y. With this type of censoring, data from the model alternative distributions tend to be less heavily censored than data from the worst case distributions.

The optimality of the RPT for the case of no nuisance parameters in the alternative distributions is established in Sections 4 and 5. In this case, the RPT exhibits a kind of finite sample minimaxity over neighborhoods of the idealized distributions, and no other nonparametric p-value has larger minimum Bahadur slope over the distributions in the alternative neighborhoods (Section 4). The RPT also offers the best compromise between robustness and optimality at the idealized alternative since no other test with the same gross-error-sensitivity has larger slope at the idealized alternative (Section 5).

The large sample optimality of the RPT with data-dependent censoring is explored in Section 6 for exponential shift models with location-scale nuisance parameters. The new test proposed in Section 3 is shown to be asymptotically optimal against restricted fixed-width neighborhoods since no other nonparametric test has a p-value with larger minimum slope over the restricted neighborhoods. RPTs with censoring based on robust estimates of the nuisance parameters, censoring based on order statistics of the combined samples, or upper and lower censoring based on the unpooled samples as proposed in Section 3 are shown to have maximin power over diminishing neighborhoods.

2. A robust two-sample permutation test. One approach to robustifying a permutation test is to replace its test statistic with a more robust statistic. For example, the difference of sample means can be replaced by the difference of sample medians, as suggested by Randles and Wolfe (1979, page 348), or by the difference of M-estimators, as suggested by Higgins and Keller (1979). An obvious question is which robust test statistic is best?

A robust test is appropriate when the relevant alternative distributions cannot be stated precisely but approximations or idealizations of these distributions can be specified. The distributions that are considered possibly relevant when an approximate model is specified determine which robust test statistic is best. The relevant distributions are defined here by drawing the following analogy with the one-sample robust testing framework developed by Huber (1965, 1968).

Take $\underline{X} = (X_1, \cdots, X_m)$ and $\underline{Y} = (Y_1, \cdots, Y_n)$ to be two random samples and let $\underline{Z} = (Z_1, \cdots, Z_N)$, $N = m + n$, be the ordered observations of the combined samples. The null hypothesis of interest is H_0: the N observations $(\underline{X}, \underline{Y})$ are exchangeable. Under all alternatives the N observations are assumed to be independent. Under the idealized *core* alternative, $X_1, \cdots, X_m$ are iid with left-continuous df F_0 and $Y_1, \cdots, Y_n$ are iid with left-continuous df G_0. Take the densities of F_0 and G_0 with respect to a measure ν to be f_0 and g_0 and assume that their log-likelihood ratio, $\ell(z) = \log(g_0(z)/f_0(z))$, is nondecreasing a.e.

As in Huber (1965, 1968), departures from the core df's F_0, G_0 are admitted by introducing the following neighborhoods:

$$\mathscr{F} = \{F\colon (1 - \varepsilon_1)F_0(t) - \delta_1 \le F(t) \le (1 - \varepsilon_1)F_0(t) + \varepsilon_1 + \delta_1 \text{ for all } t\}$$

$$\mathscr{G} = \{G\colon (1 - \varepsilon_2)G_0(t) - \delta_2 \le G(t) \le (1 - \varepsilon_2)G_0(t) + \varepsilon_2 + \delta_2 \text{ for all } t\}$$

with ε_i, $\delta_i \in [0, 1]$ sufficiently small that $\mathscr{F}$ and $\mathscr{G}$ are disjoint. The distributions in $\mathscr{F}$ are considered possibly relevant when F_0 is specified. The collection $\mathscr{F}$ contains ε_1-contaminated versions of F_0 and df's within Kolmogorov or Lévy distance δ_1 of F_0. If F_0 is stochastically smaller than G_0, then $\mathscr{F}_0$ and $\mathscr{G}_0$ may be enlarged to include all F's that are stochastically smaller than F_0 and all G's that are stochastically larger than G_0. Denote by $\mathscr{H}^N$ the set of df's of $(\underline{X}, \underline{Y})$ such that $X_1, \cdots, X_m, Y_1, \cdots, Y_n$ are independent, $X_i \sim F_i \in \mathscr{F}$, $i = 1, \cdots, m$, and $Y_j \sim G_j \in \mathscr{G}$, $j = 1, \cdots, n$. Let $\mathscr{F}^m$ be the set of m-dimensional marginal df's of $\underline{X}$ associated with $\mathscr{H}^N$ and let $\mathscr{G}^n$ be the analogous set for $\underline{Y}$.

Huber (1968) showed that the least favorable df's F_*, G_* for the one-sample test of $\mathscr{F}^m$ vs. $\mathscr{G}^m$ have the following densities f_*, g_* with respect to ν. With $w_i = \delta_i/(1 - \varepsilon_i)$ and $v_i = (\varepsilon_i + \delta_i)/(1 - \varepsilon_i)$, $i = 1, 2$,

$$(2.1)\quad f_*(z) = \begin{cases} (1 - \varepsilon_1)(v_2 f_0(z) + w_1 g_0(z))/(v_2 + w_1 e^{k_1}) & \ell(z) < k_1 \\ (1 - \varepsilon_1) f_0(z) & k_1 \le \ell(z) < k_2 \\ (1 - \varepsilon_1)(w_2 f_0(z) + v_1 g_0(z))/(v_1 e^{k_2} + w_1) & \ell(z) \ge k_2 \end{cases}.$$

$$(2.2)\quad g_*(z) = \begin{cases} (1 - \varepsilon_2)(v_2 f_0(z) + w_1 g_0(z))/(v_2 e^{-k_1} + w_1) & \ell(z) < k_1 \\ (1 - \varepsilon_2) g_0(z) & k_1 \le \ell(z) < k_2 \\ (1 - \varepsilon_2)(w_2 f_0(z) + v_1 g_0(z))/(v_1 + w_2 e^{-k_2}) & \ell(z) \ge k_2 \end{cases}$$

where the constants k_1 and k_2 are the unique solutions to

$$(2.3)\quad \begin{aligned} e^{k_1} F_0\{\ell(Z) < k_1\} - G_0\{\ell(Z) < k_1\} &= v_2 + w_1 e^{k_1} \\ e^{-k_2} G_0\{\ell(Z) > k_2\} - F_0\{\ell(Z) > k_2\} &= v_1 + w_2 e^{-k_2} \end{aligned}$$

and $F_0\{\cdot\}$ denotes the probability measure corresponding to the df $F_0(\cdot)$.

For an example, take F_0 to be the normal$(-1/2, 1)$ df and G_0 to be the normal$(1/2, 1)$ df and take $\varepsilon_1 = \varepsilon_2 = \varepsilon$, $\delta_1 = \delta_2 = \delta$. Then $\ell(z) = z$, $k_1 = k_2 \equiv k$ and k is the unique solution, if any, to $e^{-k}\Phi(-k + .5) - \Phi(-k - .5) = (\varepsilon + \delta + \delta e^{-k})/(1 - \varepsilon)$ where Φ is the normal(0, 1) df. If there is no solution for some choice of ε, δ, then $\mathscr{F}$ and $\mathscr{G}$ are not disjoint for that choice of ε, δ. If $\varepsilon = 0$ then the maximum δ for which $\mathscr{F}$ and $\mathscr{G}$ can be distinguished is .192, and if $\delta = 0$ then the maximum ε is .277. Table 1 gives the value of k for several ε, δ pairs. A blank entry signifies that $\mathscr{F}$ and $\mathscr{G}$ are not disjoint for that choice of ε, δ.

Note that the log-likelihood ratio of g_* to f_* at z is a constant plus $\ell_*(z) = \text{med}\{k_1, \ell(z), k_2\}$. Huber's one-sample maximin test of $\mathscr{F}^m$ vs. $\mathscr{G}^m$ is based on the

TABLE 1
Censoring points of the RPT statistic for normal (−.5, 1), normal (.5, 1) df's

$\delta \backslash \varepsilon$	0	.0001	.001	.01	.05	.10	.15	.20	.25
0	∞	2.93	2.31	1.58	.95	.62	.40	.23	.08
.0001	2.92	2.75	2.28	1.58	.95	.62	.40	.23	.08
.001	2.28	2.26	2.09	1.54	.94	.62	.40	.22	.07
.01	1.51	1.51	1.48	1.28	.84	.56	.35	.18	.03
.05	.82	.82	.81	.75	.53	.32	.16	.01	
.10	.44	.44	.43	.40	.25	.09			
.15	.18	.18	.17	.15	.03				

sum over the sample of ℓ_*, which is a censored version of the log-likelihood ratio for the core df's. The two-sample RPT is here defined to be based on the same statistic. Namely, the RPT statistic is $T_N(\underline{Y}) = n^{-1} \sum \ell_*(Y_i)$, or, equivalently, $n^{-1} \sum \ell_*(Y_i) - m^{-1} \sum \ell_*(X_i)$. The RPT p-value L_{*N} is evaluated by computing T_N for each sample of size n that can be drawn without replacement from $\underline{Z}$ and then calculating the proportion of the $\binom{N}{n}$ values of T_N that are at least as large as the observed $T_N(\underline{Y})$.

The calculations for the RPT are no more onerous than the calculations for any other permutation test based on a sum since all N observations are censored at the same k_1 and k_2. In particular, the RPT p-value can be evaluated using Green's algorithm. If the X's were censored at different points than the Y's, the observations would have to be recensored with each partition of $\underline{Z}$ in order to determine the support of the permutation distribution, and Green's algorithm would not apply. To reduce the "clumpiness" that might result in the permutation distribution if the censoring points k_1 and k_2 are close together and N is small, censored observations may be spread out a little near k_i rather than piled at k_i. Their original order could then be preserved.

3. A RPT with estimated location-scale. Often the core alternative distributions share a nuisance location-scale parameter $\theta = (\mu, \sigma)$ and differ by a shift parameter $\Delta > 0$. Unfortunately, the RPT of Section 2 is inappropriate for this case because it depends on the unknown θ. For example, when the core df's are normal$(\mu + .5\sigma, \sigma^2)$ and normal$(\mu - .5\sigma, \sigma^2)$ and $\varepsilon_1 = \varepsilon_2$, $\delta_1 = \delta_2$ the RPT is based on $\sum \text{med}\{\mu - K\sigma, Y_i, \mu + K\sigma\}$ where K is given in the example of Section 2. When the core densities are proportional to $\exp(-.5\sigma^{-1} | z - \mu + \Delta |)$ and $\exp(-.5\sigma^{-1} | z - \mu - \Delta |)$, the log-likelihood ratio is $\text{med}\{-\Delta, \sigma^{-1}(z - \mu), \Delta\}$ (plus a constant) and for $\varepsilon_1 = \varepsilon_2$, $\delta_1 = \delta_2$ the RPT is based on $\sum \text{med}\{\mu - \sigma \min(K, \Delta), Y_i, \mu + \sigma \min(K, \Delta)\}$ with $K = -K_1 = K_2$ defined by (2.3) for $\theta = (0, 1)$. For exponential scale core densities $\gamma \exp(-\gamma x)$ and $\gamma\delta \exp(-\gamma\delta y)$, the transformed data $\log(x)$ and $\log(y)$ have common location $-\log(\gamma) = \mu$ and a shift $-\log(\delta) = \Delta$, and a RPT for δ is based on $\sum \text{med}\{K_1 + \mu, \log(Y_i), K_2 + \mu\}$ with K_1, K_2 defined by (2.3) for $\mu = 0$. Although the RPT depends on the nuisance θ in these examples, the dependence has a simple form. The RPT statistic can be written as a sum of censored observations and the nuisance parameter θ affects only the censoring points $k_i(\theta) = \mu + K_i\sigma$, $i = 1, 2$, K_i being θ-free.

Data-dependent censoring for situations in which the nuisance parameter affects only the censoring of the RPT is illustrated in this section and is justified for large samples from exponential families in Section 6. Two types of data-dependent censoring are considered. The first is to estimate θ robustly and censor the data at $k_i(\hat{\theta}) = \hat{\mu} + K_i\hat{\sigma}$ where K_i is appropriate for $\theta = (0, 1)$. The second is to censor the data at two order statistics, and thereby circumvent the difficulties of estimating location and scale robustly for asymmetric distributions. With either type of censoring, $\hat{\underline{k}}$ need not be a symmetric function of the N observations $(\underline{X}, \underline{Y})$, but all N observations are censored only once at $\hat{\underline{k}}$ and then the RPT p-value is calculated as if the censoring had not been determined from the data.

To fix the ideas, consider the following data taken from McIneath and Cohen

(1970) as given by Box, Hunter and Hunter (1978, page 158). The data are the specific airways resistances thirty minutes after administration of a bronchodilating aerosol automatically (sample X) or by hand (sample Y).

X	11.60	11.60	13.65	17.22	8.25	6.20	41.50	6.96	8.40	9.00	5.18	3.00
Y	17.00	22.80	21.60	20.40	11.20	14.00	52.25	7.50	12.20	18.85	6.05	4.05

It seems likely that the data should be paired, but for purposes of illustration the pairing will be ignored. A boxplot shows that 41.5 is an outlier in the X sample and 52.25 is an outlier in the Y sample. For testing H_0: X and Y have the same distribution against H_1: Y is stochastically larger than X, MINITAB gives a t-test p-value of .1286, a normal scores p-value of .1695, and a Wilcoxon p-value of .0595. The p-value of the permutation test based on $\sum Y_i$ is .1430.

To develop a RPT for this example, take the core alternative df's to be F_θ: normal$(\mu - .5\sigma, \sigma^2)$ and G_θ: normal$(\mu + .5\sigma, \sigma^2)$, and take the contamination parameters to be $\varepsilon_1 = \varepsilon_2 = \delta_1 = \delta_2 = .002$. The medians of the least favorable df's $F_{*\theta}$ and $G_{*\theta}$ defined by equations (2.1)–(2.2) are then $\mu - .31\sigma$ and $\mu + .31\sigma$, respectively. As would be expected, these medians are more nearly equal than are the medians of the core df's F_θ and G_θ.

The censoring points of the RPT based on $F_{*\theta}$, $G_{*\theta}$ are $\mu \pm 1.8656\sigma$. One way to proceed is to estimate μ and σ robustly. A difficulty is that many natural estimates of μ and σ depend on whether Δ is assumed to be zero as required by H_0, or to be .5 as specified by the core df's, or to be .31 as suggested by the medians of the least favorable df's. When Δ is large, the RPT may have reasonable power even if the censoring is incorrect. But, when Δ is small, the neighborhoods F_θ and G_θ may be difficult to distinguish and the power of the RPT may be sensitive to the censoring. This reasoning suggests taking $\Delta = 0$ (Proposition 1 of Section 6 gives an asymptotic justification).

One choice of estimates for μ and σ that is appropriate when $\Delta = 0$ is the MAD estimate defined by $\hat{\sigma}_z = (2/\pi)^{1/2}\text{med}\,|Z_i - \text{med}(Z_i)|$ and the M-estimate $\hat{\mu}_z$ defined by $\sum \psi(Z_i - \hat{\mu}_z) = 0$ with $\psi(u) = \text{med}\{-1.5\hat{\sigma}_z, u, 1.5\hat{\sigma}_z\}$. For the data above, $\hat{\sigma}_z = 4.31$, $\hat{\mu}_z = 12.01$, and $\hat{\underline{k}} = (3.98, 20.05)$. Censoring all 24 observations at 3.98 and 20.05 and basing a permutation test on the average of the censored Y's gives a p-value of .0407. A second possibility is to estimate σ and μ separately for each sample, pool the estimates according to $\hat{\mu} = .5\hat{\mu}_x + .5\hat{\mu}_y$ and $\hat{\sigma} = (.5\hat{\sigma}_x^2 + .5\hat{\sigma}_y^2)^{1/2}$, and censor all 24 observations at $\hat{\mu} \pm 1.8656\hat{\sigma}$. For these data, $\hat{\mu} = 12.44$, $\hat{\sigma} = 3.51$, $\hat{\underline{k}} = (5.89, 18.98)$ and the p-value of the permutation test based on the censored data is .0377.

Another possibility is to avoid estimating μ and σ and instead choose two order statistics of the combined samples $\underline{Z}$ for $\hat{\underline{k}}$. Unfortunately, with this choice of $\hat{\underline{k}}$, data from the idealized df's F_θ, G_θ tends to be more heavily censored than data from the worst case (least favorable) df's $F_{*\theta}$, $G_{*\theta}$. To see this, take $m = n$, F_θ symmetric about $\mu - \Delta$, $G_\theta(\cdot) = F_\theta(\cdot - 2\Delta)$, $\varepsilon_1 = \varepsilon_2 > 0$ and $\delta_1 = \delta_2 = 0$. Then the RPT censors at $\mu \pm K\sigma$ for some θ-free K. Suppose $\hat{k}_1$ and $\hat{k}_2$ are the $[N\beta + 1]$ and $[N - N\beta]$ order statistics of $\underline{Z}$. Let β be the fraction of observations that should be left-censored in the worst case, i.e., $\beta = Q_{*\theta}(\mu - K\sigma)$ where $Q_{*\theta} = .5(F_{*\theta} + G_{*\theta})$ is the df of the observations of the combined samples. Note that

$Q_{*\theta}$ is symmetric about μ. For this choice of β, (2.1), (2.2) and (2.3) imply that the probability an observation is between $\mu - K\sigma$ and $\mu + K\sigma$ under the core df $.5(F_\theta + G_\theta)$ exceeds $1 - 2\beta$. Hence, under the core df, the $[N\beta + 1]$ order statistic of $\underline{Z}$ tends to be larger than $\mu - K\sigma$ and the $[N - N\beta]$ order statistic of $\underline{Z}$ tends to be smaller than $\mu + K\sigma$. Since under the worst case df's these order statistics of $\underline{Z}$ tend to $\mu \pm K\sigma$, censoring tends to be heavier when the data are not contaminated (i.e., the data are from F_θ, G_θ) than when the data are most contaminated (i.e., from $F_{*\theta}$, $G_{*\theta}$). This behavior is intuitively unreasonable, and Theorem 6 suggests that it is not optimal.

An examination of how outliers affect the permutation p-value suggests another $\hat{\underline{k}}$ that does tend to censor most heavily at the least favorable distributions. First, right tail outliers in the X's are more serious than left tail outliers in the X's for testing. Second, whether a particular X is too large, falsely dominates the Y's and should be right-censored depends on the values of the Y's. Third, the proportion of Y's to be right-censored should be controlled since larger Y's provide stronger support for the alternative. Together these considerations suggest right-censoring all N observations at the $[n\beta_2]$ ordered Y, denoted $Y_{[n\beta_2]}$. Similar arguments suggest left-censoring all N observations at the $[m\beta_1 + 1]$ ordered X, denoted $X_{[m\beta_1+1]}$, for some $\beta_1 < \beta_2$. The censoring fractions β_1 and β_2 should be appropriate for the worst case $F_{*\theta}$, $G_{*\theta}$, i.e., take $\beta_1 = F_{*\theta}(\mu + K_1\sigma)$ and $\beta_2 = G_{*\theta}(\mu + K_2\sigma)$. Note that β_1, β_2 are independent of θ. Also, the censoring is asymptotically heaviest under the least favorable df's since $F(\mu + K_1\sigma) \geq F_{*\theta}(\mu + K_1\sigma)$ for all $F \in \mathscr{F}_\theta$ and $G(\mu + K_2\sigma) \leq G_{*\theta}(\mu + K_2\sigma)$ for all $G \in \mathscr{G}_\theta$ where $\mathscr{F}_\theta$, $\mathscr{G}_\theta$ are the neighborhoods around F_θ, G_θ. The RPT with censoring at $X_{[m\beta_1+1]}$ and $Y_{[n\beta_2]}$ will be denoted $\widehat{\text{RPT}}$.

Applying the $\widehat{\text{RPT}}$ in the example above gives $\beta_1 = G_{*\theta}(\mu + 1.8656\sigma) = .0839$, $1 - \beta_2 = .0839$, $\hat{k}_1 = X_{[2]} = 5.18$, $\hat{k}_2 = Y_{[11]} = 22.80$, and the $\widehat{\text{RPT}}$ p-value is .0401. Spreading the censored observations out to 5.17, 5.16 and 22.81, 22.82 again gives a p-value of .0401.

4. Optimality of the RPT: the case of no nuisance parameters. When the core df's F_0 and G_0 are specified completely, the RPT is based on an average of log-likelihood ratios bounded by fixed constants. Clearly, the RPT is robust against outliers. In this section, we show that it is the best robust test against all departures from the core df's F_0 and G_0, that are represented by the neighborhoods $\mathscr{F}$ and $\mathscr{G}$.

Every similar (i.e., nonparametric) test of H_0: exchangeability has a discrete p-value taking values in $\{i/\binom{N}{n}: i = 1, \cdots, \binom{N}{n}\}$ (Lehmann and Stein, 1949). Here, as in Kempthorne and Folks (1971), a p-value is called optimal if it is stochastically smaller than any other p-value with the same or fewer achievable levels. This definition makes comparisons with the RPT p-value L_{*N} difficult, however, because the achievable levels of L_{*N} vary with the combined samples $\underline{Z}$. When $\underline{Z}$ requires more censoring, L_{*N} has fewer achievable levels. Typically, discrete p-values are made comparable by randomization so that each p-value assumes all values between 0 and 1. Yet L_{*N} would probably not be randomized in practice, and randomizing L_{*N} seems to complicate rather than to simplify the finite sample theory for the RPT. Therefore, rather than randomize L_{*N}, we force comparability with the RPT by randomizing all other p-values after observation of $\underline{Z}$, if necessary, so that they have the same achievable levels as L_{*N}. For example, suppose for a given $\underline{Z}$ that L_{*N} has achievable levels $\{.4, 1.\}$ and L_N has achievable levels $\{.3, .6, 1.\}$. Then the L_N test statistic assumes values s_1, s_2, s_3 with s_1 being most contradictory of H_0. If s_1 is recorded with probability ¼ and s_2 is recorded with probability ¾ whenever s_3 is recorded, then the expected p-value of the redefined test has achievable levels $\{.4, 1.\}$ and is comparable with L_{*N}.

Theorem 1 below establishes that the RPT is minimax in the following sense for finite samples. No other p-value restricted to have the same achievable levels as L_{*N} can be stochastically smaller than L_{*N} under Huber's least favorable df's F_*, G_*, and L_{*N} is stochastically larger under F_*, G_* than it is under any other pair of df's F, G in $\mathscr{F}$, $\mathscr{G}$. In this sense, the worst behavior of the RPT under $\mathscr{F}$, $\mathscr{G}$ is better than (or at least no worse than) the worst behavior of any other nonparametric test of H_0 against $\mathscr{F}$, $\mathscr{G}$.

Theorem 1 applies to the unconditional behavior of L_{*N} averaged over all the combined samples $\underline{Z}$. It would be stronger if its conclusion were valid for the conditional distribution of L_{*N} given each $\underline{Z}$, and if the support of other nonparametric p-values did not have to be modified. An asymptotic version of this stronger result is given by Theorem 2.

THEOREM 1. *The RPT p-value L_{*N} is minimax for testing H_0: exchangeability against H: $\mathscr{F}^m$, $\mathscr{G}^n$ in the following sense. Let L_N be any other nonparametric p-value and after observing $\underline{Z}$ choose a restriction L'_N of L_N to the support of L_{*N}. Then* $\sup \underline{H}\{L'_N \geq \alpha\} \geq \sup \underline{H}\{L_{*N} \geq \alpha\}$ *for every* $\alpha \in [0, 1]$ *where the supremum is taken over* $\underline{H} \in \mathscr{H}^N$.

PROOF. Conditional on $\underline{Z}$ the RPT, being based on the log-likelihood ratio of F_* and G_*, is most powerful among similar tests of H_0 against the simple alternative H_*: $X_1, \cdots, X_m \sim$ iid F_*, $Y_1, \cdots, Y_n \sim$ iid G_* (Lehmann and Stein, 1949). Thus, given $\underline{Z}$,

$$\underline{H}_*\{L_{*N} \geq \alpha \mid \underline{Z}\} \leq \underline{H}_*\{L_N \geq \alpha \mid \underline{Z}\} \tag{4.1}$$

where $\underline{H}_*$ is the joint df of $(\underline{X}, \underline{Y})$ under H_*, and it suffices to show that

$$\underline{H}_*\{L_{*N} \geq \alpha\} \geq \underline{H}\{L_{*N} \geq \alpha\} \quad \text{for every} \quad \underline{H} \in \mathscr{H}^N \quad \text{and every } \alpha. \tag{4.2}$$

For $\alpha \in (0, 1]$, choose the integer N_α satisfying $N_\alpha - 1 < \alpha\binom{N}{n} \leq N_\alpha$ and denote by $T_{(1)} \geq T_{(2)} \geq \cdots$ the $\binom{N}{n}$ ordered sample means of samples of size n taken without replacement from $(\ell_*(Z_1), \cdots, \ell_*(Z_N))$. Then $L_{*N} < \alpha$ iff $T_N(\underline{Y}) > T_{(N\alpha)}$. Equivalently, in terms of $T'_N = m^{-1} \sum \ell_*(X_i)$ and the $\binom{N}{n}$ ordered sample means $T'_{(1)} \leq T'_{(2)} \leq \cdots$ of samples of size m taken without replacement from $(\ell_*(Z_1), \cdots, \ell_*(Z_N))$, the RPT p-value satisfies $L_{*N} < \alpha$ iff $T'_N(\underline{X}) < T'_{(N\alpha)}$.

Therefore, it suffices to show that for $\underline{H} \in \mathscr{H}^N$,

(i) $\underline{H}(T_N(\underline{Y}) > T_{(N_\alpha)}) \geq \underline{H}_*(T_N(\underline{Y}) > T_{(N_\alpha)})$

(ii) $\underline{H}(T'_N(\underline{X}) < T'_{(N_\alpha)}) \geq \underline{H}_*(T'_N(\underline{X}) < T'_{(N_\alpha)})$.

The proof of (i) and (ii) relies on the least favorable character of F_*, G_* described by Huber (1968). Namely, for any $F \in \mathscr{F}$, $G \in \mathscr{G}$ and any t,

$$G\{\ell_*(Z) < t\} \leq G_*\{\ell_*(Z) < t\} \leq F_*\{\ell(Z) < t\} \leq F\{\ell(Z) < t\}, \tag{4.3}$$

and the ordering remains true if $<$ is replaced by $\leq$ throughout.

The proof of (i) proceeds by considering X_1 alone, applying (4.3), and then considering each of $X_2, \cdots, X_m$ in turn. First,

$$\underline{H}(T_N(\underline{Y}) \leq T_{(N_\alpha)}) = E_{F_2,\cdots,F_m,\underline{G}} F_1(T_N(\underline{Y}) \leq T_{(N_\alpha)} \mid X_2, \cdots, X_m, \underline{Y}).$$

For fixed $(x_2, \cdots, x_m, \underline{Y})$, $T_{(N_\alpha)}$ is increasing in x_1 and $T_N(\underline{Y})$ is constant in x_1. Therefore, by (4.3) for some $k(X_2, \cdots, X_m, \underline{Y})$, it holds that

$$\begin{aligned} F_1\{T_N(\underline{Y}) &\leq T_{(N_\alpha)} \mid X_2, \cdots, X_m, \underline{Y}\} \\ &= F_1\{\ell_*(X_1) \geq k(X_2, \cdots, X_m, \underline{Y}) \mid X_2, \cdots, X_m, \underline{Y}\} \\ &\leq F_*\{\ell_*(X_1) \geq k(X_2, \cdots, X_m, \underline{Y}) \mid X_2, \cdots, X_m, \underline{Y}\} \\ &= F_*\{T_N(\underline{Y}) \leq T_{(N_\alpha)}) \mid X_2, \cdots, X_m, \underline{Y}\}. \end{aligned}$$

That is,

$$\underline{H}(T_N(\underline{Y}) \leq T_{(N_\alpha)}) \leq E_{F_*,F_2,\cdots,F_m,\underline{G}} \mathscr{I}(T_N(\underline{Y}) \leq T_{(N_\alpha)}),$$

where $\mathscr{I}(A)$ is the indicator of the event A. The extension to (i) is straightforward. The proof of (ii) is analogous with T', Y's, and G's replacing T, X's, and F's. □

As N increases, all appropriate p-values approach zero under a fixed alternative. The exponential rate at which a p-value approaches zero under an alternative is defined by Bahadur (1967, 1971) to be the slope of the test. That is, the slope is the a.s.[F, G] limit, if it exists, of $-N^{-1}\log L_N$ where the permissible values of N have a partition m, n and $m/N \to \lambda \in (0, 1)$. (Bahadur's slope is actually twice the slope as defined here.) Typically, $-N^{-1}\log L_N$ is also asymptotically normal$(c(F, G), N^{-1}\sigma^2(F, G))$ under F, G (Lambert and Hall, 1982).

The slope $c(F, G)$ of a permutation test based on $n^{-1} \sum u(Y_i)$ at alternative df's F, G is given in Lambert and Hall (1982). When $X_1, \cdots, X_m \sim$ iid F, $Y_1, \cdots, Y_n \sim$ iid G and $u(Z)$ has finite absolute third moment for $Z \sim F$ and $Z \sim G$, the slope of the permutation test conditional on $\underline{Z}$ is

$$\begin{aligned} &c(F, G) \\ &\quad = -\lambda \int \log(\lambda b_1^{-1} e^{b_2 u(z)} + \lambda)\, dF(z) - \bar{\lambda} \int \log(\bar{\lambda} b_1 e^{-b_2 u(z)} + \bar{\lambda})\, dG(z) \end{aligned} \tag{4.4}$$

where $\bar{\lambda} = 1 - \lambda$ and $b_i \equiv b_i(F, G)$, $i = 1, 2$, are the unique solutions to

$$\text{(4.5)}\qquad \begin{aligned} \bar{\lambda} &= \int (1 + b_1 e^{-b_2 u(z)})^{-1}\, d(\lambda F(z) + \bar{\lambda} G(z)) \\ \bar{\lambda} \int u(z)\, dG(z) &= \int u(z)(1 + b_1 e^{-b_2 u(z)})^{-1}\, d(\lambda F(z) + \bar{\lambda} G(z)). \end{aligned}$$

If u is the log-likelihood ratio of F and G, then $b_1 = \lambda/\bar{\lambda}$ and $b_2 = 1$. Although the combined samples vector $\underline{Z}$ does not appear in the expression (4.4) for $c(F, G)$, the slope $c(F, G)$ does describe the conditional limiting behavior of L_N given $\underline{Z}$ (see Bahadur and Raghavachari (1970) for details). For that reason, $c(F, G)$ is sometimes called a conditional slope.

Theorem 2 states that the RPT is asymptotically optimal against $\mathscr{H}^N$ since, given almost any $\underline{Z}$, the minimum rate under $\mathscr{F}$, $\mathscr{G}$ at which a nonparametric p-value based on a sum approaches zero is never larger than the minimum rate at which L_{*N} approaches zero.

THEOREM 2. *If $m/N \to \lambda \in (0, 1)$ as $N \to \infty$, then the RPT maximizes the minimum slope under the alternative $X_1, \cdots, X_m \sim$ iid $F \in \mathscr{F}$, $Y_1, \cdots, Y_n \sim$ iid $G \in \mathscr{G}$ among all nonparametric tests having statistics of the form $\sum u(Y_i)$.*

PROOF. The two-sample permutation test based on the log-likelihood ratio for F, G has the largest slope among all tests of H_0: exchangeability against the simple alternative F, G (Bahadur and Raghavachari, 1970; Lambert and Hall, 1982). Therefore, the RPT has maximal slope against H_*: F_*, G_*. The slope of the RPT at F, G, denoted by $c_*(F, G)$, is given by (4.4), (4.5) with ℓ_* substituted for u.

Suppose for now that $c_*(F, G)$ cannot be increased by changing b_1 to $\lambda/\bar{\lambda}$ and b_2 to 1. Then

$$c_*(F, G) \geq -\lambda \int \log(\bar{\lambda} e^{\ell_*(z)} + \lambda)\, dF(z) - \bar{\lambda} \int \log(\lambda e^{-\ell_*(z)} + \bar{\lambda})\, dG(z).$$

The stochastic ordering in inequality (4.4) then implies that the right side is not increased if F, G is replaced by F_*, G_*. With the substitution, the right side becomes $c_*(F_*, G_*)$. Therefore, it only remains to show that $c_*(F, G)$ cannot be increased by changing b_1 to $\lambda/\bar{\lambda}$ and b_2 to 1. To this end, let

$$d(s, t) = -\lambda \int \log(\lambda s^{-1} e^{t\ell_*(z)} + \lambda)\, dF(z) - \bar{\lambda} \int \log(\bar{\lambda} s e^{-t\ell_*(z)} + \bar{\lambda})\, dG(z).$$

Then

$$\frac{\partial}{\partial s} d(s, t) = \lambda s^{-1} - \int (s + e^{t\ell_*(z)})^{-1}\, d(\lambda F(z) + \bar{\lambda} G(z))$$

$$\frac{\partial}{\partial t} d(s, t) = \bar{\lambda} \int \ell_*(z)\, dG(z) - \int \ell_*(z)(s e^{-t\ell_*(z)} + 1)^{-1}\, d(\lambda F(z) + \bar{\lambda} G(z))$$

and the pair of equations (4.5) imply that $(\partial/\partial s) d(s, t) = (\partial/\partial t) d(s, t) = 0$ at b_1,

b_2. Since $d(b_1, b_2) > 0$ and $d(s, t) \to -\infty$ as $(s, t) \to (0, \pm\infty)$ or $(\infty, \pm\infty)$, it follows that $d(s, t)$ is maximized by b_1, b_2. □

5. Optimality of the RPT at the model F_0, G_0. Theorems 1 and 2 describe the optimality of the RPT for testing exchangeability against *neighborhoods* of the core df's F_0, G_0. They do not disallow the possibility that some other equally robust test dominates the RPT at F_0, G_0. Theorem 3 disallows this possibility if test robustness is measured by the bounds on the influence functions of the test; i.e., by gross-error-sensitivity (see Hampel, 1974).

Influence functions of unconditional and conditional p-values are defined and interpreted in Lambert (1981). By analogy with Hampel's influence functions of estimators, the influence functions $\Omega_i(\cdot, F, G)$, $i = 1, 2$, of a two-sample test under an alternative F, G are defined for the a.s. limit of $-N^{-1}\log L_N$, i.e., for the slope $c(F, G)$. Specifically,

$$\Omega_1(x, F, G) = \lim_{\eta\downarrow 0}\eta^{-1}[c((1-\eta)F + \eta\delta_x, G) - c(F, G)]$$

$$\Omega_2(y, F, G) = \lim_{\eta\downarrow 0}\eta^{-1}[c(F, (1-\eta)G + \eta\delta_y) - c(F, G)],$$

if the limits exist, where δ_z is the distribution supported on z. With $X_1, \cdots, X_m$ being called the first sample and $Y_1, \cdots, Y_n$ being called the second, the ith sample influence function at z, F, G indicates how an observation of z in the ith sample affects the p-value when the remaining X_i's and Y_j's are assumed to be iid F and iid G and $m/N \sim \lambda$. Bounded influence functions are associated with (a) insensitivity to a few outliers or inliers, and continuous influence functions are associated with (b) insensitivity to many small errors, such as those introduced by rounding, and with (c) insensitivity to slight departures from the assumed models F, G.

For a permutation test based on $n^{-1}\sum u(Y_i)$,

$$\Omega_{u1}(x, F, G) = -\lambda\log(1 + b_1^{-1}e^{b_2u(x)}) + \lambda\int\log(1 + b_1^{-1}e^{b_2u(z)})\,dF(z)$$

(5.1)

$$\Omega_{u2}(y, F, G) = -\bar{\lambda}\log(1 + b_1e^{-b_2u(y)}) + \bar{\lambda}\int\log(1 + b_1e^{-b_2u(z)})\,dG(z)$$

where b_1, b_2 are the solutions to the pair of equations (4.5). These influence functions are bounded above and continuous in x and y for any nondecreasing continuous function u; they are bounded above and below and continuous in F and G only if u is bounded and continuous. Therefore, a permutation test p-value based on $\sum u(Y_i)$ cannot be pulled arbitrarily close to zero, but it is not robust in the sense of properties (a)–(c) unless u is bounded and continuous. Therefore, rank tests with unbounded score functions are not robust, the log-likelihood ratio permutation test is not necessarily robust, and the RPT is robust in the sense of properties (a)–(c).

Theorem 3 states that the RPT is the optimal robust test of exchangeability against F_0, G_0 since no other permutation test with a test statistic of the form $n^{-1}\sum u(Y_i)$ has both smaller gross-error-sensitivity and larger slope at F_0, G_0. Because the influence functions of all these permutation tests are bounded above, gross-error-sensitivity is here defined to be the lower bound on the influence function. A similar concept of optimality at a model for robust unconditional tests is defined in terms of the influence function of the test statistic and asymptotic power of the test in Ronchetti (1982) and Rousseeuw (1982). For unconditional tests, which depend on the data only through the test statistic, the two approaches coincide.

THEOREM 3. *Denote the slope and influence functions of the RPT by* c_*, Ω_{*1}, *and* Ω_{*2} *respectively. A permutation test based on a statistic* $\sum u(Y_i)$ *with* $\inf_z \Omega_{ui}(z, F_0, G_0) \geq \inf_z \Omega_{*i}(z, F_0, G_0)$, *for* $i = 1, 2$, *has slope* $c_u(F_0, G_0)$ *no larger than* $c_*(F_0, G_0)$.

PROOF. Since all slopes and influence functions are evaluated at F_0, G_0, the arguments F_0, G_0 will be suppressed. Denote $\inf_z \Omega_{*i}(z)$ by ω_i. First consider the problem of choosing a function u to maximize the slope c_u with no constraints on the influence functions. Let c_{u,b_1,b_2} be defined by (4.4) for arbitrary $b_1 > 0$, b_2. The argument used in the proof of Theorem 2 shows that c_u (with b_1, b_2 defined by (4.5)) satisfies

$$
\begin{aligned}
c_u &= \sup_{b_1,b_2} c_{u,b_1,b_2} \\
&= \sup - \int (\lambda \log(1 + b_1^{-1} e^{b_2 u(z)}) f_0(z) + \bar{\lambda} \log(1 + b_1 e^{-b_2 u(z)}) g_0(z))\, d\nu(z) \\
&\quad - \lambda \log \lambda - \bar{\lambda} \log \bar{\lambda}.
\end{aligned}
$$

The maximum over u of c_{u,b_1,b_2} is attained by choosing $u(z)$ pointwise according to $u_0(z) = b_2^{-1}(\ell(z) + \log(\bar{\lambda} b_1/\lambda))$. All the permutation tests generated by considering any $b_1 > 0$, $-\infty < b_2 < \infty$ have the same slope, and $\sup_u c_u = c_{u_0}$ for any $b_1 > 0$, b_2.

Now impose the constraints $\Omega_{ui}(z) \geq \omega_i$, $i = 1, 2$. Then the slope c_u is maximized by choosing $u_0(z)$ as $u_0(z) = b_{02}^{-1}(\ell(z) + \log(\bar{\lambda} b_{01}/\lambda))$ for any b_{01}, b_{02} whenever

$$
\Omega_{u_0 1}(z) = -\lambda \log(1 + \bar{\lambda} e^{\ell(z)}/\lambda) + \lambda \int \log(1 + b_{01}^{-1} e^{b_{02} u_0(t)})\, dF_0(t) \geq \omega_1
$$

$$
\Omega_{u_0 2}(z) = -\bar{\lambda} \log(1 + \lambda e^{-\ell(z)}/\bar{\lambda}) + \bar{\lambda} \int \log(1 + b_{01} e^{-b_{02} u_0(t)})\, dG_0(t) \geq \omega_2.
$$

If the first constraint on $\Omega_{u_0 1}$ is violated for some z' then it is violated for every larger z, and if it is satisfied for some z'' then it is satisfied for every smaller z since ℓ is nondecreasing. A similar statement holds for the constraint on $\Omega_{u_0 2}$. Hence, there are z_1 and z_2 such that

$$
\begin{aligned}
u_0(z) &= b_{02}^{-1}(\ell(z) + \log(\bar{\lambda} b_{01}/\lambda)) && \text{if } z \in (z_1, z_2) \\
u_0(z) &\in [u_0(z_1), u_0(z_2)] && \text{if otherwise.}
\end{aligned}
$$

Choose and fix $z < z_1$. Then $u_0(z)$ equals $u_0(z_1)$ if the integrand in c_u is smaller with $u_0(z)$ set equal to $u_0(z_1)$ than it would be for $u_0(z)$ set equal to any other

$u_0(z'), z' \in (z_1, z_2)$. Specifically, $u_0(z) = u_0(z_1)$ if

$$\bar{\lambda}\log(1 + e^{-\ell(z_1)})g_0(z) + \lambda\log(1 + e^{\ell(z_1)})f_0(z) \le \bar{\lambda}\log(1 + e^{-\ell(z')})g_0(z) + \lambda\log(1 + e^{\ell(z')})f_0(z),$$

or, equivalently, if

$$\left(1 + \frac{\bar{\lambda}g_0(z)}{\lambda f_0(z)}\right)\log\left(\frac{1 + e^{-\ell(z_1)}}{1 + e^{-\ell(z')}}\right) \le \ell(z') + \ell(z_1).$$

The latter inequality is satisfied if $z = z_1$, since c_u is maximized by choosing $u_0(z_1)$ at $z = z_1$. Since the left side is nondecreasing in z and the right is independent of z, the inequality must be satisfied for every $z < z_1$. That is, $u_0(z) = u_0(z_1)$ for $z < z_1$. Likewise, if $z > z_2$ then $u_0(z) = u_0(z_2)$ if for every $z' < z_2$

$$\left(1 + \frac{\lambda f_0(z)}{\bar{\lambda}g_0(z)}\right)\log\left(\frac{1 + e^{\ell(z_2)}}{1 + e^{\ell(z')}}\right) \le \ell(z_2) - \ell(z').$$

Again, because $z = z_2$ satisfies this inequality, any $z > z_2$ must also satisfy it, and $u_0(z) = u_0(z_2)$ for $z > z_2$. Therefore, $u_0(z) = b_{02}^{-1}(\text{med}\{\ell(z_1), \ell(z), \ell(z_2)\} + \log(\bar{\lambda}b_0/\lambda))$. Since $u_0(z)$ is a censored log-likelihood ratio and the corresponding permutation test has the same bounds on its influence functions as the RPT, the permutation test based on $u_0(z)$ must be equivalent to the RPT. □

6. Optimality of a RPT for shift with estimated location-scale. The robust testing alternative can be generalized to include a shared nuisance parameter $\theta \in \Theta$ as follows:

$$H: X_1, \cdots, X_m \sim \text{iid } F, \quad Y_1, \cdots, Y_n \sim \text{iid } G, \quad (F, G) \in \cup_\theta(\mathscr{F}_\theta, \mathscr{G}_\theta) \equiv \cup_\theta \mathscr{H}_\theta$$

$$\mathscr{F}_\theta = \{F: (1 - \varepsilon_1)F_\theta(t) - \delta_1 \le F(t) \le (1 - \varepsilon_1)F_\theta(t) + \varepsilon_1 + \delta_1 \text{ for all } t\}$$

$$\mathscr{G}_\theta = \{G: (1 - \varepsilon_2)G_\theta(t) - \delta_2 \le G(t) \le (1 - \varepsilon_2)G_\theta(t) + \varepsilon_2 + \delta_2 \text{ for all } t\}$$

for ε_i, δ_i independent of θ, $i = 1, 2$. The set of relevant joint df's of $(\underline{X}, \underline{Y})$ corresponding to a given θ will be denoted $\mathscr{H}_\theta^N = (\mathscr{F}_\theta^m, \mathscr{G}_\theta^n)$ with the understanding that within samples the observations are identically distributed. Only location-scale nuisance parameters θ are considered here.

In many examples the RPT depends on θ only through the censoring $\underline{k}(\theta)$. Three such examples were given in Section 3. Many others can be constructed by considering location-scale exponential family core df's for tests of shift. That is, for nuisance parameter $\theta = (\mu, \sigma)$ and shift $\Delta > 0$, define $F_\theta = P_{\theta,-\Delta}$ and $G_\theta = P_{\theta,\Delta}$ where $P_{\theta,\Delta}$ has density $p_{\theta,\Delta}$ satisfying $p_{\theta,\Delta}(z)/p_{\theta,0}(z) = a(\theta, \Delta)\exp(\Delta\sigma^{-1}t(x - \mu))$ for a θ-free function t. The RPT statistic with this choice of F_θ, G_θ is $\sum \text{med}\{\mu + \sigma K_1, t(Y_i), \mu + \sigma K_2\}$ for a θ-free $\underline{K}$ depending on $\underline{\varepsilon}$, $\underline{\delta}$. Without loss of generality take $t(x) = x$. Denote the permutation test based on $\sum \text{med}\{b_1, Y_i, b_2\}$ by RPT($\underline{b}$), and denote the censoring points $(\mu + (\sigma/\Delta)K_1, \mu + (\sigma/\Delta)K_2)$ by $\underline{k}(\theta)$.

The asymptotic behavior of a RPT with data-dependent censoring $\hat{\underline{k}}$ is investigated below. Throughout we assume $m/N \to \lambda \in (0, 1)$. First, the asymptotic power of the RPT($\hat{\underline{k}}$) is investigated for several kinds of data-dependent censoring in the robust testing framework of Huber-Carol (1970) and Rieder (1978, 1981). In this framework the shift Δ approaches zero as N increases, forcing the contamination parameters $\underline{\varepsilon}$, $\underline{\delta}$ also to approach zero to preserve disjointedness of $\mathscr{F}_\theta$, $\mathscr{G}_\theta$. Conditions that ensure a RPT($\hat{\underline{k}}$) has asymptotic maximin power against such diminishing neighborhoods are given in Theorem 4. Asymptotic maximin power is achieved when censoring is based on either $\hat{\mu} + \hat{\sigma}\underline{K}$ for robust $\hat{\mu}$, $\hat{\sigma}$, or on two order statistics of $\underline{Z}$, or on an order statistic of $\underline{X}$ and an order statistic of $\underline{Y}$ as in the $\widehat{\text{RPT}}$.

Finally, the asymptotic behavior of the RPT($\hat{\underline{k}}$) against fixed neighborhoods is explored. A RPT($\hat{\underline{k}}$) is said to have maximin slope if for each θ no other test has a larger minimum slope under $\mathscr{F}_\theta$, $\mathscr{G}_\theta$. The optimality of the $\widehat{\text{RPT}}$ in this sense is established as follows. Conditions under which a RPT($\hat{\underline{k}}$) has the same slope under alternative df's F, G as a permutation test with data-free censoring $\underline{k}(F, G)$ are given in Theorem 5. Since $\hat{\underline{k}}$ may have different limits under different F, G, the permutation test with data-free censoring that is equivalent in slope to the RPT($\hat{\underline{k}}$) varies with F, G. A condition on the equivalent censoring $\underline{k}(F, G)$ that ensures the RPT($\hat{\underline{k}}$) achieves its minimum slope over F_θ, G_θ at Huber's least favorable $F_{*\theta}$, $G_{*\theta}$ is then given in Theorem 6. In part, the condition requires that if $\varepsilon_i' \le \varepsilon_i$ and $\delta_i' \le \delta_i$, $i = 1, 2$, then data from F, G in the $\underline{\varepsilon}'$, $\underline{\delta}'$ neighborhood should be no more heavily censored than data from an F, G outside the $(\underline{\varepsilon}', \underline{\delta}')$-neighborhood but inside the $(\underline{\varepsilon}, \underline{\delta})$-neighborhood. Of all the tests discussed in Section 3, to date only the RPT($\hat{\underline{k}}$) has been shown to have the latter property, and only if the neighborhoods $\mathscr{F}_\theta$, $\mathscr{G}_\theta$ are somewhat restricted. An exact statement for the $\widehat{\text{RPT}}$ is given as Theorem 7.

6.1 *Asymptotic power of a RPT*($\hat{\underline{k}}$). Rieder (1978) defines neighborhoods $\mathscr{F}_{\theta N}$, $\mathscr{G}_{\theta N}$, that are appropriate for studying the asymptotic power of robust tests. These neighborhoods involve a shift $\Delta_N = N^{-1/2}\Delta$, core df's $F_{\theta N} = P_{\theta,-\Delta_N}$, $G_{\theta N} = P_{\theta,\Delta_N}$, and contamination parameters $\varepsilon_{iN} = N^{-1/2}\varepsilon_i$, $\delta_{iN} = N^{-1/2}\delta_i$. Wang (1981) develops a robust $C(\alpha)$ test for the parametric hypotheses $\mathscr{F}_{\theta N}$ vs $\mathscr{G}_{\theta N}$ $\theta \in \Theta$ and shows that the test has asymptotic maximin power under certain conditions on the estimator of θ. A variant of his condition C, given as (*) in Theorem 4, guarantees that the RPT($\hat{\underline{k}}$) has asymptotic maximin power over $\cup_\theta(\mathscr{F}_{\theta N}, \mathscr{G}_{\theta N})$.

THEOREM 4. *Suppose $X_1, \cdots, X_m \sim$ iid F_N and $Y_1, \cdots, Y_n \sim$ iid G_N for some $F_N \in F_{\theta N}$, $G_N \in G_{\theta N}$, $\theta \in \Theta$. Also suppose for each θ,*

$$(*)\qquad \begin{aligned} &\text{given}\quad \eta > 0 \quad \text{there exist } e \text{ and } N_0 \text{ such that for} \quad N \ge N_0 \\ &\underline{H}_N\{N^{1/2} \mid \hat{k}_{iN} - k_{iN}(\theta) \mid \le e\} \ge 1 - \eta \quad i = 1, 2 \end{aligned}$$

where $\underline{H}_N$ is the joint df of $(\underline{X}, \underline{Y})$. Then the RPT($\hat{\underline{k}}_N$) has asymptotic maximin power against the sequence $\{\cup_\theta(\mathscr{F}_{\theta N}, \mathscr{G}_{\theta N})\}$.

PROOF. Choose and fix θ and df's $F_N \in \mathscr{F}_{\theta N}$, $G_N \in \mathscr{G}_{\theta N}$. Denote the joint df of $(\underline{X}, \underline{Y})$ under F_N, G_N by $\underline{H}_N$. Note that $k_{iN}(\theta) = \mu + C_{iN}\sigma/\Delta_N$ for some θ-free $\underline{C}_N$, and that C_{iN}/Δ_N is uniformly bounded (Rieder, 1978). To simplify the notation, suppress θ. With $u^* = \text{med}\{k_{iN}, u, k_{2N}\}$ and $\hat{u} = \text{med}\{\hat{k}_{1N}, u, \hat{k}_{2N}\}$, the RPT($\underline{k}_N$) statistic can be written as $T_N^* = (mn/N)^{1/2}(n^{-1} \sum Y_i^* - m^{-1} \sum X_i^*)/S_N^*$ where S_N^* is the sample standard deviation of $(\underline{X}^*, \underline{Y}^*)$ and the RPT($\hat{\underline{k}}_N$) statistic can be written as $\hat{T}_N = (mn/N)^{1/2}(n^{-1} \sum \hat{Y}_i - m^{-1} \sum \hat{X}_i)/\hat{S}_n$ where $\hat{S}_N$ is the sample standard deviation of the N values $(\hat{X}, \hat{Y})$. Because of (*) and the contiguity of the distributions in $\{\mathscr{F}_N, \mathscr{G}_N\}$,

$$(n^{-1} \sum Y_i^* - m^{-1} \sum X_i^*) - (n^{-1} \sum \hat{Y}_i - m^{-1} \sum \hat{X}_i) = o_p(N^{-1/2}) \quad \text{under} \quad \{\underline{H}_N\}.$$

Therefore, following Hoeffding (1952, conditions A′, B and Theorem 6A), the RPT($\hat{\underline{k}}_N$) and the RPT($\underline{k}_N$) have the same limiting power whenever $R_N^* = \max_i N^{-1/2} \,|\, Z_i^* - N^{-1} \sum Z_i^* \,|\, /S_N^* = o_p(1)$ under $\{\underline{H}_N\}$. That $R_N^* = o_p(1)$ under $\{\underline{H}_N\}$ holds since Z_i^* is bounded.

The limiting power of the RPT($\underline{k}_N$) under $\{F_N, G_N\}$ depends on the limiting unconditional distribution of T_N^*/S_N^*. Since $\sup_N |\, Z_N^* \,| < \infty$, as in Rieder (1978) there is a $\sigma^* > 0$ such that for every choice of $(F_N, G_N) \in (\mathscr{F}_N, \mathscr{G}_N)$, $T_N^* - \hat{T}_N - (\mu_N(\underline{H}_N)/\sigma^*)$ is asymptotically normal(0, 1) for $\mu_N(\underline{H}_N) = E_{H_N}(T_N^*)$. Therefore, when the RPT($\underline{k}_N$) has size α, for large N its power differs negligibly from $\Phi(\Phi^{-1}(\alpha) + \mu_N/\sigma^*)$ (Hoeffding, 1952). Since $\inf_{\mathscr{H}^N} \mu_N(\underline{H}_N) \equiv \mu_N^*$ occurs at the least favorable (F_{*N}, G_{*N}) and since no other size α test has a power function with an infimum over $\mathscr{H}^N$ larger than $\Phi(\Phi^{-1}(\alpha) + \mu_N^*/\sigma^*)$ for large N (Rieder, 1978), the asymptotic minimaxity of the RPT($\underline{k}_N$) and the RPT($\hat{\underline{k}}$) is now established. □

Many different $\hat{\underline{k}}_N$ satisfy the sufficient condition (*). For the case σ known, Proposition 1 gives a set of conditions under which an M-estimator of location satisfies (*). It is stated for $\hat{\mu}$ based on the combined samples; it also holds for pooled M-estimates of μ. Its proof, which proceeds along the lines of the proof of Lemma 4 of Huber (1964), is omitted. Note that Proposition 1 implies that the asymptotic power of the RPT is preserved when θ is estimated as if Δ_N were zero, as was done in Section 3.

PROPOSITION 1. *Assume $\sigma = 1$. Define $\hat{\mu}_N$ by $\sum \psi(Z_i - \hat{\mu}_N) = 0$ for a monotone increasing, bounded, uniformly continuous function ψ. Define $\zeta(u, Q)$ to be $\int \psi(z + u)\, dQ(z)$. Suppose $\zeta(0, P_0) = 0$ and $\partial\zeta(u, P_0)/\partial u$ is strictly positive and continuous in a neighborhood of zero where P_0 corresponds to $\theta = (0, 1)$, $\Delta = 0$. Then the RPT($\hat{\underline{k}}$) with censoring at $\hat{k}_{iN} = \Delta_N^{-1} K_{iN} + \hat{\mu}_N$ has asymptotic maximin power against $\{\cup_\theta(\mathscr{F}_{\theta N}, \mathscr{G}_{\theta N})\}$.*

Proposition 2 gives conditions which imply that the $\widehat{\text{RPT}}$ has asymptotic maximin power. The same conditions also imply that the RPT($\hat{\underline{k}}$) with censoring at order statistics of the combined samples is asymptotically maximin. The proof of Proposition 2 is a minor modification of the proof of Lemma 1 in Jaeckel (1971) and is omitted.

PROPOSITION 2. *Restrict $\mathscr{F}_{\theta N}$ to df's that have a density over the intervals $(k_{iN}(\theta) - \gamma, k_{iN}(\theta) + \gamma)$, $i = 1, 2$, for some $\gamma > 0$. Further suppose that the density is bounded below by a positive constant on these intervals. Denote the restricted*

$\mathscr{F}_{\theta N}$ by $\mathscr{F}^c_{\theta N}$. Construct $\mathscr{G}^c_{\theta N}$ analogously. Then the $\widehat{\text{RPT}}$ has asymptotic maximin power against $\{\cup_\theta(\mathscr{F}^c_{\theta N}, \mathscr{G}^c_{\theta N})\}$.

6.2 *The slope of a RPT($\hat{k}$).* Conditions under which a RPT($\hat{k}$) has the same slope under an alternative F, G as a permutation test with data-free censoring are given in Theorem 5.

THEOREM 5. *Take $X_1, \cdots, X_m$ to be iid $Q_{1\theta}$ and $Y_1, \cdots, Y_n$ to be iid $Q_{2\theta}$, and suppose $m/N \to \lambda \in (0, 1)$. Also suppose the following conditions are met.*

(6.1) *For each θ, $T_N(\theta) \equiv T_N(\theta; \underline{X}, \underline{Y})$ converges a.s. $(Q_{1\theta}, Q_{2\theta})$ to a data-free $T(\theta)$.*

(6.2) *For each θ there is a neighborhood $\mathscr{I}_\theta$ of $T(\theta)$ and a continuous function c_θ on $\mathscr{I}_\theta$ satisfying $-N^{-1}\log Q_N\{T_N(\theta) \geq t \mid \underline{Z}\} - c_\theta(t) = o(1)$ a.s. $(Q_{1\theta}, Q_{2\theta})$ on $\mathscr{I}_\theta$ where $Q_N\{\cdot \mid \underline{Z}\}$ denotes the permutation probability distribution.*

(6.3) *There is a data-free real-valued function $j(\cdot, \cdot)$ such that $|T_N(\theta_1) - T_N(\theta_2)| \leq j(\theta_1, \theta_2)$ and $j(\theta_1, \theta_2) = O(\theta_1 - \theta_2)$.*

(6.4) *The estimator $\hat{\theta}_N \equiv \hat{\theta}_N(\underline{X}, \underline{Y})$ converges to θ a.s. $(Q_{1\theta}, Q_{2\theta})$.*

Let $L_N(\hat{\theta}_N)$ be the permutation p-value calculated by first evaluating $\hat{\theta}_N$ and then determining the permutation distribution of $T_N(\hat{\theta}_N)$ with $\hat{\theta}_N$ held fixed for all partitions. Then for every θ, $L_N(\theta_N)$ and $L_N(\hat{\theta}_N)$ have slope $c_\theta(T(\theta))$ at $Q_{1\theta}$, $Q_{2\theta}$.

PROOF. Theorem 7.2 of Bahadur (1971) implies that $L_N(\theta)$ has slope $c_\theta(T(\theta))$ at $Q_{1\theta}$, $Q_{2\theta}$ and that $L_N(\hat{\theta}_N)$ has slope $c_\theta(T(\theta))$ if (i) $T_N(\hat{\theta}_N) \to T(\theta)$ a.s. and (ii) $-N^{-1}\log Q_N\{T_N(\hat{\theta}_N) \geq t \mid \underline{Z}\} - c_\theta(t) \to 0$ a.s. for t in a neighborhood of $T(\theta)$. Since $j(\hat{\theta}_N, \theta) \to 0$ a.s., (i) holds. To establish (ii), fix sequences $x_1, x_2, \cdots; y_1, y_2, \cdots$ such that $\hat{\theta}_N \to \theta$ and $-N^{-1}\log Q_N\{T_N(\theta) \geq t \mid \underline{Z}\} \to c_\theta(t)$ on $\mathscr{I}_\theta$. Note that by (6.3), $Q_N\{T_N(\hat{\theta}_N) \geq t \mid \underline{Z}\} \leq Q_N\{T_N(\theta) \geq t - j(\hat{\theta}_N, \theta) \mid \underline{Z}\}$. Since $\hat{\theta}_N$ is held fixed under permutations of $\underline{Z}$ and there is an N sufficiently large that $j(\hat{\theta}_N, \theta) < \eta$, the right side is no larger than $Q_N\{T_N(\theta) \geq t - \eta \mid \underline{Z}\}$. Taking η sufficiently small such that $(T(\theta) - 2\eta, T(\theta) + 2\eta) \subset \mathscr{I}_\theta$ gives $Q_N\{T_N(\hat{\theta}_N) \geq t \mid \underline{Z}\} \leq \exp(-Nc_\theta(t - \eta) + o(N))$. Similarly, $Q_N\{T_N(\hat{\theta}_N) \geq t \mid \underline{Z}\} \geq \exp(-Nc_\theta(t + \eta) + o(N))$. Letting $\eta \to 0$ completes the proof. □

It is straightforward to show that the RPT statistic with censoring at $\underline{k}(\theta)$ satisfies condition (6.3). If $|k_i(\theta)| < \infty$, then $T_N(\theta)$ has finite absolute third moment and (6.2) is satisfied (Lambert and Hall, 1982). Consequently, if the censoring points are not reevaluated for each partition of $\underline{Z}$, then the RPT($\hat{\underline{k}}$) has the same slope as the RPT($\underline{k}$) when $\hat{\underline{k}} \to \underline{k}$ a.s.

As discussed at the beginning of the section, the limit of $\hat{\underline{k}}$ may vary with the underlying df's F, G. Theorem 6 states that if censoring is asymptotically most severe at the least favorable df's $F_{*\theta}$, $G_{*\theta}$ and asymptotically correct at $F_{*\theta}$, $G_{*\theta}$ in the sense that $\hat{\underline{k}} \to \underline{k}(\theta)$ under $F_{*\theta}$, $G_{*\theta}$ then the slope of the RPT($\hat{\underline{k}}$) under $\mathscr{F}_\theta$, $\mathscr{G}_\theta$ is minimized at $F_{*\theta}$, $G_{*\theta}$. Recall that in Section 3, the censoring proportions of the $\widehat{\text{RPT}}$ were based on $F_{*\theta}$, $G_{*\theta}$. To simplify the notation in Theorem 6 and its proof, the dependence on θ is suppressed.

THEOREM 6. *Let F belong to an $(\varepsilon_1', \delta_1')$-neighborhood of the core df F_0 and let G belong to an $(\varepsilon_2', \delta_2')$-neighborhood of the core df G_0 for some $0 \le \varepsilon_i' \le \varepsilon_i$, $0 \le \delta_i' \le \delta_i$, $i = 1, 2$. Denote the censoring points of the optimal RPT for $\underset{\sim}{\varepsilon}$, $\underset{\sim}{\delta}$ ($\underset{\sim}{\varepsilon}'$, $\underset{\sim}{\delta}'$, resp.) by $\underset{\sim}{k}$ ($\underset{\sim}{k}'$, resp.). If $\hat{\underset{\sim}{k}} \to k$ a.s. (F_*, G_*) and $\hat{\underset{\sim}{k}} \to \underset{\sim}{t}$ a.s. (F, G) where $k_1' \le t_1 \le k_1$ and $k_2 \le t_2 \le k_2'$, then the slope of the RPT$(\hat{\underset{\sim}{k}})$ is no smaller under F, G than it is under F_*, G_*.*

PROOF. The RPT$(\hat{\underset{\sim}{k}})$ under F_*, G_* has slope

$$-\bar{\lambda} \int \log(\bar{\lambda} + \lambda e^{-z_*})\, dG_*(z) - \lambda \int \log(\lambda + \bar{\lambda} e^{z_*})\, dF_*(z) - \lambda \log \lambda - \bar{\lambda} \log \bar{\lambda}$$

where $z_* = \operatorname{med}\{k_1, z, k_2\}$. The stochastic ordering (4.3) implies that this slope is no larger than $J(k_1, k_2) - \lambda \log \lambda - \bar{\lambda} \log \bar{\lambda}$ where

$$\begin{aligned} J(t_1, t_2) = & -\bar{\lambda} \int_{t_1}^{t_2} \log(\bar{\lambda} + \lambda e^{-z})\, dG(z) - \lambda \int_{t_1}^{t_2} \log(\lambda + \bar{\lambda} e^{z})\, dF(z) \\ & - \bar{\lambda} \log(\bar{\lambda} + \lambda e^{-t_2}) \bar{G}(t_2) - \lambda \log(\lambda + \bar{\lambda} e^{t_2}) \bar{F}(t_2) \\ & - \bar{\lambda} \log(\bar{\lambda} + \lambda e^{-t_1}) G(t_1) - \lambda \log(\lambda + \bar{\lambda} e^{t_1}) F(t_1). \end{aligned}$$

Because

$$\frac{\partial J(t_1, t_2)}{\partial t_1} = \frac{\lambda \bar{\lambda}(G(t_1) - \exp(t_1) F(t_1))}{\bar{\lambda} \exp(t_1) + \lambda}$$

and

$$\frac{\partial J(t_1, t_2)}{\partial t_2} = \frac{\lambda \bar{\lambda}(\bar{G}(t_2) - \exp(t_2) \bar{F}(t_2))}{\bar{\lambda} \exp(t_2) + \lambda},$$

it follows that $J(t_1, t_2)$ is (i) nondecreasing at t_2 if $\bar{G}(t_2) > \exp(t_2)\bar{F}(t_2)$ and (ii) nonincreasing at t_1 if $G(t_1) < \exp(t_1) F(t_1)$.

For any t_2, $\bar{G}(t_2) \ge (1 - \varepsilon_2')\bar{G}_0(t_2) - \delta_2'$ and $\bar{F}(t_2) \le (1 - \varepsilon_1')\bar{F}_0(t_2) + \delta_1' + \varepsilon_1'$. If $\varepsilon_1' < \varepsilon_2'$ then $\bar{F}(t_2) < (1 - \varepsilon_2')\bar{F}_0(t_2) + \delta_1' + \varepsilon_2'$. Therefore, if (i) holds for $\varepsilon_1' \ge \varepsilon_2'$, (i) also holds for $\varepsilon_1' < \varepsilon_2'$. When $\varepsilon_1' \ge \varepsilon_2'$,

$$\bar{G}(t_2) - e^{t_2}\bar{F}(t_2) \ge (1 - \varepsilon_2')\left(\bar{G}_0(t) - \frac{\delta_2'}{1 - \varepsilon_2'} - e^{t_2}\bar{F}_0(t_2) - \frac{\delta_1' + \varepsilon_1'}{1 - \varepsilon_1'} e^{t_2}\right).$$

Since for $k_2 < t_2 < k_2'$ there are $\delta_i'' \ge \delta_i'$ and $\varepsilon_i'' \ge \varepsilon_i'$ $(i = 1, 2)$ such that

$$e^{-t_2}\bar{G}_0(t_2) - \bar{F}_0(t_2) = \frac{\varepsilon_1'' + \delta_1''}{1 - \varepsilon_1''} + \frac{\delta_2''}{1 - \varepsilon_2''} e^{-t_2}$$

(see (2.1)), (i) does hold when $\varepsilon_1' \ge \varepsilon_2$. The proof of (ii) is similar. Finally, the slope of the RPT$(\hat{\underset{\sim}{k}})$ at F, G would be $J(t_1, t_2) - \lambda \log \lambda - \bar{\lambda} \log \bar{\lambda}$ if the constants b_1, b_2 determined by (4.5) were $\lambda/\bar{\lambda}$ and 1. Since the solutions to (4.5) maximize the slope as a function of b_1, b_2, the slope of the RPT$(\hat{\underset{\sim}{k}})$ at F, G is no smaller than $J(t_1, t_2) - \lambda \log \lambda - \bar{\lambda} \log \bar{\lambda}$. □

Although the $\widehat{\text{RPT}}$ censors most heavily at $F_{*\theta}$, $G_{*\theta}$ asymptotically, Theorem 6 does not imply that the $\widehat{\text{RPT}}$ has maximin slope over $\mathscr{F}_\theta$, $\mathscr{G}_\theta$ because the inequalities $F^{-1}(\beta_1) \geq k_1'(\theta)$ and $G^{-1}(\beta_2) \leq k_2'(\theta)$ may be violated for some F, G in an $(\underset{\sim}{\varepsilon}', \underset{\sim}{\delta}')$-neighborhood of F_θ, G_θ. (These inequalities seem unnecessarily strong, but no weaker sufficient condition has been found.) The question of whether the $\widehat{\text{RPT}}$ has maximin slope over all of $U_\theta(\mathscr{F}_\theta, \mathscr{G}_\theta)$ is thus unanswered. A smaller, but nondiminishing neighborhood over which the $\widehat{\text{RPT}}$ does have maximin slope is described in Theorem 7. Strong consistency of order statistics, which is required in Theorem 7, is proved in Wellner (1977).

THEOREM 7. *Set* $\hat{k}_1 = X_{(m\beta_1+1)}$ *and* $\hat{k}_2 = Y_{(n\beta_2)}$ *with* $\beta_1 = F_{*\theta}(k_1(\theta))$, $\beta_2 = G_{*\theta}(k_2(\theta))$. *If* $\hat{\underset{\sim}{k}} \to \underset{\sim}{k}(\theta)$ *a.s.* $(F_{*\theta}, G_{*\theta})$ *then the* $\widehat{\text{RPT}}$ *has maximin slope against all df's F, G that satisfy the following three conditions*:

(6.5) $\hat{\underset{\sim}{k}} \to (F^{-1}(\beta_1), G^{-1}(\beta_2))$ *a.s.* (F, G)

(6.6) *F, G belong to an* $(\underset{\sim}{\varepsilon}', \underset{\sim}{\delta}')$*-neighborhood of* F_θ, G_θ *for some* $\varepsilon_i' \leq \varepsilon_i$, $\delta_i' \leq \delta_i$

(6.7) $(1 - \varepsilon_1')F_\theta(k_1'(\theta)) + \varepsilon_1' + \delta_1' \leq \beta_1$ *and* $(1 - \varepsilon_2')G_\theta(k_2'(\theta)) - \delta_2' \geq \beta_2$.

If $\delta_1 = \delta_2 = 0$ *and either* $\varepsilon_1 = 0$ *or* $\varepsilon_2 = 0$, *then the* $\widehat{\text{RPT}}$ *has maximin slope against the df's* $(F, G) \in \cup_\theta(\mathscr{F}_\theta, \mathscr{G}_\theta)$ *for which* (6.5) *holds.*

REMARK. There always exist $\varepsilon_i' \leq \varepsilon_i$ and $\delta_i' \leq \delta_i$ for which (6.7) holds. To see this, note that $\partial k_i/\partial\varepsilon_j$, $j = 1, 2$, are nonnegative for $i = 1$ and nonpositive for $i = 2$, and that $(1 - \varepsilon_1)F_\theta(k_1(\theta)) + \varepsilon_1 + \delta_1$ is nondecreasing and $(1 - \varepsilon_2)G_\theta(k_2(\theta)) - \delta_2$ is nonincreasing in $\varepsilon_1, \varepsilon_2, \delta_1, \delta_2$.

PROOF OF THEOREM 7. Let $\underset{\sim}{\varepsilon}'$, $\underset{\sim}{\delta}'$ satisfy (6.7). Because of Theorems 4 and 5 it suffices to show $F^{-1}(\beta_1) \geq k_1'(\theta)$ and $G^{-1}(\beta_2) \geq k_2'(\theta)$ for (F, G) in an $(\underset{\sim}{\varepsilon}', \underset{\sim}{\delta}')$-neighborhood of (F_θ, G_θ). For such F, G, $F(k_1'(\theta)) \leq (1 - \varepsilon_1')F_\theta(k_1'(\theta)) + \varepsilon_1' + \delta_1' \leq \beta_1$ and similarly $G(k_2'(\theta)) \geq \beta_2$, as was to be shown. Finally, if $\varepsilon_1 = \delta_1 = \delta_2 = 0$, then $k_2 = \infty$ and $G(k_2(\theta)) = 1 = \beta_2$ for all proper G. Also, $F_\theta^{-1}(\beta_1) \leq k_1'(\theta)$ for all $\varepsilon_2' \leq \varepsilon_2$ since $k_1'(\theta)$ is decreasing in ε_1'. □

In summary, many kinds of data-dependent censoring lead to a RPT($\hat{\underset{\sim}{k}}$) with asymptotic maximum power. Yet the $\widehat{\text{RPT}}$ may nevertheless be preferable. Its advantages are asymptotic optimality over certain fixed neighborhoods, censoring behavior that is intuitively reasonable, and simplicity. Any RPT($\hat{\underset{\sim}{k}}$) can be inverted using Green's algorithm to obtain a robust estimate of shift in the presence of a shared nuisance location-scale parameter. The behavior of such estimates is currently under investigation. Finally, there is a robust version of the usual one-sample permutation test for symmetry about zero; this test can also be inverted to estimate a location parameter.

Acknowledgements. The problem of robust permutation tests was proposed to me by Professor W. J. Hall. Preliminary results on this topic were included in my thesis, which was written under his supervision. I am grateful to him for his guidance on my thesis, for sharing his insight into permutation tests and robustness with me, and for his careful reading of this manuscript.

REFERENCES

BAHADUR, R. R. (1967). Rate of convergence of estimates and test statistics. *Ann. Math. Statist.* **38** 303–324.

BAHADUR, R. R. (1971). *Some Limit Theorems in Statistics.* SIAM, Philadelphia.

BAHADUR, R. R., CHANDRA, T. P. and LAMBERT, D. (1982). Some further properties of likelihood ratios on general sample spaces. *Proc. Indian Statist. Inst. Golden Jubilee in Statist.*

BAHADUR, R. R. and RAGHAVACHARI, M. (1970). Some asymptotic properties of likelihood ratios on general sample spaces. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* 129–152, Univ. California Press.

BICKEL, P. M. and VAN ZWET, W. R. (1978). Asymptotic expansions for the power of distribution free tests in the two-sample problem. *Ann. Statist.* **6** 987–1004.

BOX, G. E. P. and ANDERSEN, S. L. (1955). Permutation theory in the derivation of robust criteria and the study of departures from assumptions. *J. Roy. Statist. Soc. Ser. B* **17** 1–34.

BOX, G. E. P., HUNTER, W. G. and HUNTER, J. S. (1978). *Statistics for Experimenters.* Wiley, New York.

EDGINGTON, E. S. (1980). *Randomization Tests.* Dekker, New York.

EFRON, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans.* SIAM, Philadelphia.

FRASER, D. A. S. (1952). *Nonparametric Methods in Statistics.* Wiley, New York.

GABRIEL, K. R. and HALL, W. J. (1983). Re-randomization inference on regression and shift effects: computationally feasible methods. *J. Amer. Statist. Assoc.* **78** 827–836.

GREEN, B. F. (1977). A practical interactive program for randomization tests of location. *Amer. Statist.* **31** 37–39.

HAMPEL, F. R. (1974). The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc.* **69** 383–393.

HIGGINS, J. J. and KELLER, S. A. (1979). Robust permutation tests. Talk presented at the 1979 Joint Statistical Meetings.

HOEFFDING, W. (1952). Large sample power of tests based on permutations of observations. *Ann. Math. Statist.* **23** 169–192.

HUBER, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35** 73–101.

HUBER, P. J. (1965). A robust version of the probability ratio test. *Ann. Math. Statist.* **36** 1753–1758.

HUBER, P. J. (1968). Robust confidence limits. *Z. Wahrsch. verw. Gebiete* **10** 269–278.

HUBER-CAROL, C. (1970). Étude asymptotique des tests robustes. These, ETH Zurich.

JAECKEL, L. A. (1971). Robust estimates of location: symmetry and asymmetric contamination. *Ann. Math. Statist.* **42** 1020–1034.

KELLER, S. A. (1979). Robust permutation tests. Master's thesis, University of South Florida.

KEMPTHORNE, D. (1952). *The Design and Analysis of Experiments.* Wiley, New York.

KEMPTHORNE, D. and FOLKS, L. (1971). *Probability, Statistics and Data Analysis.* Iowa State University Press, Ames.

LAMBERT, D. (1978). *P*-values: asymptotics and robustness. Ph.D. thesis, University of Rochester.

LAMBERT, D. (1981). Influence functions for testing. *J. Amer. Statist. Assoc.* **76** 649–657.

LAMBERT, D. (1982). Qualitative robustness of tests. *J. Amer. Statist. Assoc.* **77** 352–357.

LAMBERT, D. and HALL, W. J. (1982). Asymptotic lognormality of *p*-values. *Ann. Statist.* **10** 44–64. (Correction to asymptotic lognormality of *p*-values **11** page 348.)

LEHMANN, E. L. and STEIN, C. (1949). On the theory of some nonparametric hypotheses. *Ann. Math. Statist* **20** 28–45.

MCILNEATH, F. J and COHEN, B. M. (1970) *J. Med.* **1** 229.

ODEN, A. and WEDEL, H. (1975). An argument for Fisher's permutation test. *Ann. Statist.* **3** 518–520.

RANDLES, R. H. and WOLFE, D. A. (1979) *Introduction to the Theory of Nonparametric Statistics.* Wiley, New York.

ROUSSEEUW, P. (1982). Estimation and testing by means of optimally robust statistics. *Revue Belge de Statistique, d'Informatique et de Recherche Operationnelle* **22** 3–14.

RONCHETTI, E. (1982). Robust testing in linear models: the infinitesimal approach. Ph.D. Dissertation, Swiss Federal Institute of Technology, Zurich.

RIEDER, H. (1978). A robust asymptotic testing model. *Ann. Statist.* **6** 1080–1094.

RIEDER, H. (1981). Robustness of one- and two-sample rank tests under gross errors. *Ann. Statist.* **9** 245–265.

TUKEY, J. W., BRILLINGER, D. R. and JONES, L. V. (1978). *The Management of Weather Resources Volume II: The Role of Statistics.* Weather Modification Advisory Board, Statistical Task Force, Washington, D.C.: U.S. Government Printing Office.

VADIVELOO, J. (1983). On the theory of modified randomization tests for nonparametric hypotheses. *Comm. Statist. A—Theory Methods* **12** 1581–1596.

WANG, P. C. C. (1981). Robust asymptotic tests of statistical hypotheses involving nuisance parameters. *Ann. Statist.* **5** 1096–1106.

WELLNER, J. A. (1977). A Glivenko-Cantelli theorem and strong laws of large numbers for functions of order statistics. *Ann. Statist.* **5** 473–480.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CHICAGO
CHICAGO, ILLINOIS 60637