ON RANDOMIZED RANK SCORE PROCEDURES OF BELL AND DOKSUM[1]

By Kumar Jogdeo

Courant Institute of Mathematical Sciences, New York University

0. Introduction and summary. Bell and Doksum (1965) proposed a new class of nonparametric tests which have the advantage of possessing exact and well tabulated distributions under the null hypothesis. The basic departure from the usual nonparametric tests, suggested by the above authors is that of taking an additional sample from a known distribution such as normal, uniform, exponential, etc. and use those *observations* as the rank scores. Throughout this paper such procedures are called 'Randomized Rank Score' (RRS) as opposed to the usual Rank Score (RS) procedures.

Bell and Doksum (1965) have shown that the RRS tests have the same Pitman efficiency behaviour as the corresponding RS tests. One could interpret this result by saying that the effect of the superimposed noise of the additional sample dissipates near the null hypothesis as the sample size tends to infinity. However, as will be shown here, for the finite sample sizes the noise does create undesirable properties for the RRS tests. Firstly, in most of the familiar testing problems the power of the RRS test remains bounded away from unity as the parameter varies over the entire region of the alternative. Secondly, the confidence sets based on RRS procedure have the following peculiar property. With positive probability, the confidence set becomes the whole parameter space and the procedure completely disregards the observations of the experiment.

1. Nonresolving tests and consequences. Let $\mathcal{P} = \{P_\theta : \theta \varepsilon \Omega\}$ be a family of probability distributions defined over the sample space $\{\mathfrak{X}, \mathfrak{A}\}$ of a random variable X. A statistical test based on X is proposed to test the hypothesis H_0: $\theta \varepsilon \omega$ against the alternative H_1: $\theta \varepsilon \Omega - \omega$. Let $B \varepsilon \mathfrak{A}$ be a level α critical region, i.e

$$\sup_{\theta\varepsilon\omega} P_\theta[X \varepsilon B] = \alpha, \tag{1.1}$$

and B' be the acceptance region, $B \cup B' = \mathfrak{X}$.

Definition 1. The test based on a critical region B is resolving if

$$\sup_{\theta\varepsilon\Omega-\omega} P_\theta[X \varepsilon B] = 1; \tag{1.2}$$

it is called nonresolving if (1.2) does not hold or equivalently if

$$\inf_{\theta\varepsilon\Omega-\omega} P_\theta[X \varepsilon B'] > 0. \tag{1.3}$$

The interpretation of this notion in the familiar testing problems is that the power of a test increases to 1 as the alternative recedes from the null hypothesis. Thus a 'resolving' test distinguishes or resolves a distant alternative from the

Received 8 June 1966; revised 20 July 1966.

[1] Research supported by Ford Foundation grant for probability and statistics.

hypothesis with a high probability. This is certainly a desirable property of a test.

In view of the concept of 'Relative Distant Efficiency' (RDE), which measures the rate at which the power approaches unity as the alternative recedes from the hypothesis, it can be said that a nonresolving test will always have RDE zero when compared with a resolving one.

In the following a brief formulation of the RRS tests proposed by Bell, Doksum (1965) is given for the specific problems viz. (i) two sample problem, (ii) c-sample problem and (iii) problem of testing independence in bivariate populations.

(i) Suppose $X_1, \cdots, X_m$; $Y_1, \cdots, Y_n$ are two random samples from populations having absolutely continuous distribution functions $F(x)$ and $F(x - \Delta)$ respectively. Let $(R_1, \cdots, R_m)$ and $(S_1, \cdots, S_n)$ be the sets of ranks of X and Y observations respectively among the combined sample of $N = m + n$ observations. In order to carry out the RRS test for testing H_0: $\Delta = 0$ against H_1: $\Delta > 0$, an additional sample of N observations is taken from a standard normal population and the order statistic $Z(1) < \cdots < Z(N)$ is formed. The level α RRS test rejects H_0: $\Delta = 0$ in favor of $\Delta > 0$ whenever

$$(mn)^{\frac{1}{2}}\{n^{-1}\textstyle\sum_{i=1}^{n} Z(S_i) - m^{-1}\sum_{j=1}^{m} Z(R_j)\} > N^{\frac{1}{2}}K(\alpha), \tag{1.4}$$

where

$$(2\pi)^{-\frac{1}{2}}\textstyle\int_{K(\alpha)}^{\infty} \exp(-x^2/2)\, dx = \alpha. \tag{1.5}$$

For the two sided test of H_0: $\Delta = 0$ against $\Delta \neq 0$, the acceptance region of the level α RRS test is given by

$$-N^{\frac{1}{2}}K(\alpha/2) < (mn)^{\frac{1}{2}}\{n^{-1}\textstyle\sum_{i=1}^{n} Z(S_i) - m^{-1}\sum_{j=1}^{m} Z(R_j)\} < N^{\frac{1}{2}}K(\alpha/2). \tag{1.6}$$

(ii) For considering the c-sample problem, let X_{ij}, $j = 1, \cdots, n_i$; $i = 1, \cdots, c$, be c random samples from populations with absolutely continuous distributions $F_1, \cdots, F_c$ and R_{ij} be the rank of the jth observation in the ith sample when ranking is done among all $n_1 + n_2 + \cdots + n_c = n$ observations. Let $Z(1) < \cdots < Z(n)$ denote the ordered observations in a random sample of n observations from a standard normal population. Further let

$$\begin{aligned} Z_{i\cdot} &= \textstyle\sum_{j=1}^{n_i} Z(R_{ij})/n_i, \qquad i = 1, \cdots, c, \\ Z_{\cdot\cdot} &= \textstyle\sum_i \sum_j Z(R_{ij})/n. \end{aligned} \tag{1.7}$$

The RRS test rejects H_0: $F_1 = \cdots = F_c$ whenever

$$\textstyle\sum_{i=1}^{c} n_i(Z_{i\cdot} - Z_{\cdot\cdot})^2 > C(\alpha) \tag{1.8}$$

where $C(\alpha)$ is the upper $100\alpha\%$ cutoff point of the χ_{c-1}^2 distribution.

(iii) Let (U_i, V_i), $i = 1, \cdots, n$ be a random sample of pairs from a bivariate population with absolutely continuous distribution function $F(u, v)$. The problem is to test the hypothesis H_0: U and V are independent against the linear alternative

$$U = \theta W_1 + (1 - \theta)W_2 \tag{1.9}$$
$$V = \theta W_1 + (1 - \theta)W_3$$

where $0 < \theta \leqq 1$, and W_1, W_2 and W_3 are mutually independent random variables. Here $\theta = 0$ is equivalent to the hypothesis of independence. Such alternatives were considered by Bhuchongkul (1964). She showed that the bivariate analogue of the normal score test for the above problem has the same efficiency behaviour as that of the normal scores test in the two sample problem.

Let R_i be the rank of U_i among the U observations and S_i be that of V_i among the V observations. Two independent random samples each of size n are drawn from the standard normal population and the corresponding order statistics $Z(1) < \cdots < Z(n)$ and $Z'(1) < \cdots < Z'(n)$ are formed. Then the level α RRS test rejects the hypothesis of independence in favor of the linear alternative given by (1.9) whenever

$$\sum_{i=1}^{n} Z(R_i)Z'(S_i) > C_\rho(\alpha), \tag{1.10}$$

where $C_\rho(\alpha)$ is the upper $100\alpha\%$ cutoff point of the distribution of the sample correlation coefficient ρ, the sample being of size n and drawn from a standard spherical bivarate normal distribution.

Although not given by Bell and Doksum (1965), by using their method and results of Bhuchongkul (1964) it can be shown that the Pitman efficiency of the RRS test given by (1.10) is at least as good as the correlation test under all possible disributions of U and V.

THEOREM 1. (a) *For the problems* (i), (ii) *and* (iii) *considered above the* RRS *tests are all nonresolving for every finite sample size and at levels of significance,* $0 < \alpha < 1$, (*except for the one sided test in the two sample problem where* α *has to be less than* $\frac{1}{2}$).

b) *For the same problems sign tests, Wilcoxon tests and normal scores tests are resolving for* α *larger than the smallest natural level.*

The proof of the theorem follows from the following lemma which gives a sufficient condition for a test to be nonresolving.

Suppose $Z_{11} < Z_{12} < \cdots < Z_{1m_1}$; $Z_{21} < \cdots < Z_{2m_2}$; $\cdots$; $Z_{k1} < \cdots < Z_{km_k}$ are k ordered samples derived from k independent random samples obtained from a population having an absolutely continuous distribution function. Further let $m = m_1 + m_2 + \cdots + m_k$. Consider a discrete random variable X, independent of the Z's, which takes values on the set S of all possible m-tuples obtained by permuting the subscripts of the Z_{ij} within k groups. Thus S contains $m_1!\, m_2! \cdots m_k!$ elements. The distribution of X is assumed to depend on a parameter $\theta \in \Omega$. Let $Z(x)$ denote the m-tuple obtained from $Z_{11}, \cdots, Z_{km_k}$ by permuting the subscripts according to $x \in S$ and $T[Z(X)]$ be a real valued statistic whose distribution will depend on the parameter θ. For testing the hypothesis H_0: $\theta \varepsilon \omega$ against $\theta \varepsilon \Omega - \omega$ let A be the acceptance region of a level α test so that if $T \varepsilon A$ the hypothesis H_0 is accepted.

LEMMA 1. *The level α test based on the acceptance region for the statistic T given above is nonresolving if for every $x \varepsilon S$*

$$P[T(Z(x)) \varepsilon A] > 0. \tag{1.11}$$

PROOF. First note that due to independence of X and the Z's the probability in (1.11) does not depend on θ. Since the set S has a finite number of elements, (1.11) implies that

$$\epsilon = \min_{x \varepsilon S} P[T(Z(x)) \varepsilon A] > 0 \tag{1.12}$$

which in turn implies that for every $\theta \varepsilon \Omega$ the power of the test

$$\begin{aligned} 1 - P_\theta[T(Z(X)) \varepsilon A] & \\ = 1 - \textstyle\sum_{x \varepsilon S} P[T(Z(x)) \varepsilon A] \cdot P_\theta[X = x] & \\ \leqq 1 - \min_{x \varepsilon S} P[T(Z(x)) \varepsilon A] \textstyle\sum_{x \varepsilon S} P_\theta[X = x] & \\ = 1 - \epsilon. & \end{aligned} \tag{1.13}$$

Thus the power being bounded away from unity the test is nonresolving.

COROLLARY. *Condition* (1.11) *is satisfied if*

$$P[\textstyle\bigcap_{x \varepsilon S} \{T(Z(x)) \varepsilon A\}] > 0. \tag{1.14}$$

As a consequence, (1.14) *implies that the test based on A is nonresolving.*

PROOF. Obvious.

PROOF OF THEOREM 1. The proofs will be given for the one and two sided two sample problems only. The others follow exactly the same pattern.

The acceptance region for the two sided test is given by (1.6). The observations of the Z sample may satisfy the following inequalities simultaneously

$$\begin{aligned} (mn)^{\frac{1}{2}}\{n^{-1}\textstyle\sum_{i=1}^{n} Z(m+i) - m^{-1}\sum_{j=1}^{m} Z(j)\} &< N^{\frac{1}{2}}K(\alpha/2), \\ (mn)^{\frac{1}{2}}\{n^{-1}\textstyle\sum_{i=1}^{n} Z(i) - m^{-1}\sum_{j=1}^{m} Z(n+j)\} &> -N^{\frac{1}{2}}K(\alpha/2). \end{aligned} \tag{1.15}$$

Obviously the probability that (1.15) holds is positive and under such circumstances no matter what the X and Y observations are the RRS test will always accept the hypothesis. The Corollary to Lemma 1 applies and the test is nonresolving.

To see this happen in the case of the one sided alternative note that $\alpha < \frac{1}{2}$ implies that $K(\alpha)$ is positive. Thus, with positive probability

$$(mn)^{\frac{1}{2}}\{n^{-1}\textstyle\sum_{i=1}^{n} Z(m+i) - m^{-1}\sum_{j=1}^{m} Z(j)\} < N^{\frac{1}{2}}K(\alpha), \tag{1.16}$$

and the Corollary to Lemma 1 applies. The proof for the c-sample RRS test is similar. In considering the problem of testing independence in bivariate population the RRS test needs two samples. The results of Lemma 1 or its corollary can be applied with $k = 2$ and $m_1 = m_2 = n$.

The claim (b) of the Theorem 1 can be easily checked. The assumption of α being larger than the minimum natural level implies that there exists a point in the critical region where no randomization is needed. For example, in the case of the two sample problem $R_i < S_j$ for $i = 1, \cdots, m$ and $j = 1, \cdots, n$, will be in the critical region. As the shift parameter increases the probability of the realization of this point approaches unity. For other problems a similar argument validates (b) of the Theorem 1. This concludes the proof.

In the following we consider confidence sets based on the RRS procedure. Let $\mathfrak{X}$ be the sample space of a random observable X, usually vector valued and x be a typical member of $\mathfrak{X}$. Let the subsets $\{B(\theta_0): \theta_0 \varepsilon \Omega\}$ represent a class of acceptance regions for level α tests of the hypotheses $H(\theta_0): \theta = \theta_0$. It is well known that (see Lehmann (1959), Theorem 4, Chapter 3)

$$CS(x) = \{\theta: x \,\varepsilon\, B(\theta), \theta \,\varepsilon\, \Omega\} \tag{1.17}$$

constitutes a confidence set with confidence coefficient $1 - \alpha$.

The usual technique to generate a class of acceptance regions can be formally stated by using the folowing definition.

Definition 2. A class of acceptance regions $\{B_\theta : \theta \,\varepsilon\, \Omega\}$ is said to be "generated by $B(\theta_0)$ and U" if there exists a transformation $U: \Omega \times \mathfrak{X} \rightarrow \mathfrak{X}$ such that for every $\theta \,\varepsilon\, \Omega$ fixed, U is measurable and

$$x \,\varepsilon\, B(\theta) \Leftrightarrow U(\theta, x) \,\varepsilon\, B(\theta_0). \tag{1.18}$$

Illustrations. For testing $\Delta = \Delta_0$, $-\infty < \Delta_0 < \infty$, in the two sample problem, the acceptance regions based on the Student's t-test are generated by the one corresponding to $\Delta = 0$. The transformation used is that of shift, operated on the Y observations. The same holds for Wilcoxon acceptance regions, a fact that has been used by Lehmann (1963) to set up confidence intervals.

In the above discussion if $T(X)$ is a real valued test statistic then the acceptance regions $\{A(\theta), \theta \,\varepsilon\, \Omega\}$ could be expressed as subsets of the real line while

$$CS(t) = \{\theta: t \,\varepsilon\, A(\theta)\} \tag{1.19}$$

would be the confidence set whenever $T(X) = t$.

Theorem 2. *With the same notation as in Lemma* 1, *assume that* (1.14) *is satisfied for* $\omega = \{\theta_0\}$, $A = A(\theta_0)$ *and that the family* $\{A(\theta), \theta \,\varepsilon\, \Omega\}$ *is generated by the* RRS *acceptance region* $A(\theta_0)$. *Then, there exists a set of* Z *observations with positive probability on which the confidence set*

$$CS(t) = \Omega, \tag{1.20}$$

no matter what X *has been observed.*

Proof. Recall that the random variable X is assumed to take values in the finite set S. Consider

$$I = \{z: \textstyle\bigcap_{x \varepsilon S} [T(z(x)) \,\varepsilon\, A(\theta_0)]\}, \tag{1.21}$$

so that

$$(1.22) \qquad z \varepsilon I \Rightarrow T(z(x)) \varepsilon A(\theta_0), \qquad \text{for every } x \varepsilon S.$$

Suppose $U:\Omega \times S \rightarrow S$ is the transformation used for generating $\{A(\theta): \theta \varepsilon \Omega\}$. Then irrespective of what X is observed, since for every θ $U(\theta, X) \varepsilon S$ the relation (1.22) gives

$$(1.23) \qquad \begin{aligned} z \varepsilon I &\Rightarrow T(z(u)) \varepsilon A(\theta_0) \\ &\Rightarrow T(z(x)) \varepsilon A(\theta), \quad \text{for every } \theta \varepsilon \Omega \text{ and } x \varepsilon S. \end{aligned}$$

Consequently,

$$(1.24) \qquad \begin{aligned} z \varepsilon I \Rightarrow CS(t) &= \{\theta: T(z(x)) \varepsilon A(\theta)\} \\ &= \Omega \qquad \text{for ever } x \varepsilon S, \end{aligned}$$

and (1.14) further states that $P(I) > 0$. This completes the proof of the theorem.

Example. If the inequalities (1.15) hold, no matter what Δ_0 is added to or subtracted from Y observations the hypothesis $\Delta = 0$ is always accepted and the confidence interval based on RRS procedure becomes the whole real line disregarding what X or Y observations are.

2. Some remarks.

Remark I. Theorems 1 and 2 reveal certain unpleasant properties of the RRS tests. Of course, this unpleasantness may arise with very small probability.

The nonresolving nature of the RRS test can be used for guessing how the power function behaves for the alternatives not very close to the hypothesis, as much as Pitman efficiency is viewed as a guide to power function for large values of the sample sizes and nearby alternatives.

Remark II. While applying the RRS tests two statisticians may arrive at different conclusions with the same experimental data. One may argue that this happens with the usual tests, however, such an occurrence takes place only at the boundary of the critical region and can be considered as a genuine indifference regarding two decisions and unlike the RRS procedure it is not a commonplace.

As is customary with many applied statisticians, reporting the probability of the first kind of error is of some importance. This is useful for a scientist who is planning a similar experiment. However, such a reporting with a new test is meaningless since so much noise has been superimposed by the second sample.

Remark III. The only advantage of the new test that has been pointed out in [1] is that of being able to carry out the test with known tables. However, the recent upsurge in the study of nonparametric methods is yielding more tables for various tests which are accessible to many users. Thus it is not all that convincing that taking a random sample from the standard normal population is really time and labor saving Further the suggested device is not useful if the samples are taken from uniform or logistic distribution since the tables for the distribution of the corresponding sample means are not readily available.

3. Acknowledgment. In concluding this paper the author wants to express sincere thanks to Prof. P. K. Sen for a useful discussion.

REFERENCES

[1] BELL, C. B. and DOKSUM, K. (1965). Some new distribution-free tests. *Ann. Math. Statist.* **36** 201–214.

[2] BHUCHONGKUL, S. (1964). A class of nonparametric tests for independence in bivariate populations. *Ann. Math. Statist.* **35** 138–149.

[3] LEHMANN, E. L. (1959). *Testing Statistical Hypotheses*. Wiley, New York.

[4] LEHMANN, E. L. (1963). Nonparametric confidence intervals for a shift parameter. *Ann. Math. Statist.* **34** 1507–1512.