DISTRIBUTION-FREE TESTS OF INDEPENDENCE[1]

BY C. B. BELL[2] AND K. A. DOKSUM[3]

Institut de Statistique de l'Université de Paris

1. Introduction and summary. The object of this article is to characterize the family of all distribution-free (DF) tests of independence, and those subfamilies which are optimal for specified alternative classes. It is found (Theorem 3.2) that each DF statistic is a function of a Pitman (or permutation) statistic; and (Theorem 3.4) that the rank statistics are those whose distributions depend appropriately on the maximal invariant [Lehmann (1959), p. 227]. For parametric alternatives the MP (most powerful) [Lehmann and Stein (1949)] and the locally MP tests are found to be Pitman tests based on the likelihood function (Lemma 4.1 and Theorem 4.1); while the corresponding optimal rank tests are analogous (Lemmas 4.2, 4.3) to those in the 2-sample case [e.g. Capon (1961)], and are closely related to those of Bhuchongkul (1964). In Section 5 randomized statistics similar to those of [4] are shown (Corollary 5.2) to be asymptotically equivalent to the optimal nonrandomized statistic for specified parametric alternatives. For one reasonable nonparametric class of alternatives one proves (Theorem 6.1) that the normal scores test is minimax; while for the other class an unexpected statistic (Theorem 6.2) is minimax. Finally, in Section 7 the ideas of monotone tests developed by Chapman (1958) and others are extended, and analogous results (Theorem 7.1) are obtained for minimum power.

2. Terminology, notation and preliminaries. The study of DF tests for the independence hypothesis is essentially the study of similar sets and similar test functions. The basic ideas of Pitman (1937 a, b); (1938); Scheffé (1943); Lehmann and Stein (1949); and Bell and Bellot (1965) are generalized to construct the desired sets and functions in terms of permutation groups and related functions. For the rank tests, which occupy a central position both in theory and in practice, one needs a group of transformations which (among other things) generates the null hypothesis class. To this end one needs the following notation:

(A) *Sample.* The generic data point is $(x_1, y_1), \cdots, (x_n, y_n)$. However, for the sequel it will be more convenient to represent this point of R_{2n} as

$$z = (x, y) = (x_1, \cdots, x_n; y_1, \cdots, y_n) = (z_1, \cdots, z_{2n}).$$

(B) *Hypothesis Classes.* Ω will denote the generic class of probability measures on some R_k; more specifically one has: $\Omega_2(R_k)$, the class of continuous distributions on R_k; and $\Omega(H_0)$, the null hypothesis class $= \{F^{(n)} \cdot G^{(n)}: F, G \,\varepsilon\, \Omega_2(R_1)\}$ where

Received 6 June 1966; revised 29 October 1966.

[1] Partially supported by the National Science Foundation through NSFG-25220.

[2] Now at Case Institute of Technology.

[3] Now at University of California, Berkeley.

$(F^{(n)} \cdot G^{(n)})(z) = \prod_1^n F(x_i) \prod_1^n G(y_j);$

and $\Omega(H_0 \cup H_1) = \{H^{(n)} : H \varepsilon \Omega_2(R_2)\}$,

where $H^{(n)}(z) = \prod_1^n H(x_i, y_i)$.

DEFINITION 2.1. Ω is called NS-*complete* if, whenever v is NS (Definition 2.3) and $\int v\, dH = 0$ for all H in Ω, $P\{v \neq 0 \mid H\} = 0$ for all H in Ω.

It will be seen (Lemma 3.2) that $\Omega(H_0)$ has this property.

(C) *Groups.* $\mathcal{S}_n = \{\pi\}$, the symmetric group of the $n!$ permutations of $\{1, \cdots, n\}$; $\mathcal{S} = \mathcal{S}_n \times \mathcal{S}_n = \{\gamma = (\pi, \pi') : \pi, \pi' \varepsilon \mathcal{S}_n\}$, the direct product of $\mathcal{S}_n$ with itself; $\mathcal{S}(z) = \{\gamma(z) : \gamma \varepsilon \mathcal{S}\}$ will denote the *z-orbit.*

For the generic point on an orbit it will be convenient to write

$$\gamma(z) = [\pi(x), \pi'(y)] = [\pi(x_1), \cdots, \pi(x_n); \pi'(y_1), \cdots \pi'(y_n)].$$

It is easily seen that for the orbits:

(2.1) each element of $\Omega(H_0)$ is invariant over each $\mathcal{S}(z)$; and

(2.2) $\mathcal{S}(z)$ contains $(n!)^2$ elements for a.a. $[\Omega(H_0)]$ z.

$\mathcal{G}_2 = \{g\}$, the group of 1–1 strictly increasing transformations of R_1 onto R_1. $\mathcal{G} = \mathcal{G}_2(n) \times \mathcal{G}_2(n) = \{(g^{(n)}, f^{(n)}) : g, f \varepsilon \mathcal{G}_2\}$, where $(g^{(n)}, f^{(n)})(z) = [g(x_1), \cdots, g(x_n); f(y_1), \cdots, f(y_n)]$.

One shows easily that

(2.3) $\mathcal{G}$ is a group of 1–1 transformations of R_{2n} onto R_{2n}; and

(2.4) $\mathcal{G}$ generates $\Omega(H_0)$ in the sense that for each $H^{(n)}$ in

$$\Omega(H_0), \Omega(H_0) = \{H^{(n)}(g^{(n)}, f^{(n)}) : g, f \varepsilon \mathcal{G}_2\}.$$

(D) *Functions.* As usual φ is a test function if it is measurable and $0 \leqq \varphi \leqq 1$.

DEFINITION 2.2. A test function φ is

(i) *similar of size* α wrt $\Omega(H_0)$ if $\int \varphi\, dH = \alpha$ for all $H \varepsilon \Omega(H_0)$;

(ii) an *α-Lehmann-Stein function* if $\sum \varphi(\gamma(z)) = (n!)^2\alpha$ for a.a. $[\Omega(H_0)]$ z, where the summation is over γ in $\mathcal{S}$; and

(iii) a *classical test function* if φ is the indicator function I_A of some set A (called the critical region).

In order to construct the sets corresponding to the functions above and to emphasize the NS (nonsequential) character of the tests, one introduces the following definitions:

DEFINITION 2.3. A measurable function v is

(i) a *B-Pitman function* if $P\{v(z) = v(\gamma(z)) \mid H\} = 0$ for all H in $\Omega(H_0)$ and all γ other than the identity;

(ii) an NS *function* if $v(x, y) = v(\pi(x), \pi(y))$ for a.a. (x, y) and all π;

(iii) a BNS-*Pitman function* if v is NS and

$$P\{v(x, y) = v(\pi(x), \pi'(y)) \mid H\} = 0$$

for all H in $\Omega(H_0)$ whenever $\pi \neq \pi'$; and

(iv) a *ranking function* if it is constant on each B-rank set. (See Definition 2.5.)

(E) *Sets.* It will be seen that the important sets are also related to the orbits. To this end one defines

DEFINITION 2.4. A measurable set A is

(i) *similar of size* α wrt $\Omega(H_0)$ if $P\{A \mid H\} = \alpha$ for all H in $\Omega(H_0)$;

(ii) an *α-Scheffé set* if A contains exactly $(n!)^2\alpha$ points of a.a. orbit;

(iii) NS if for all π in $\mathcal{S}_n$ and (x, y) in A, A contains $[\pi(x), \pi(y)]$;

(iv) *almost invariant* wrt $\mathcal{G}$ and $\Omega(H_0 \cup H_1)$ if

$$P\{(A - g(A)) \cup (g(A) - A) \mid H\} = 0$$

for all $g \in \mathcal{G}$ and all H in $\Omega(H_0 \cup H_1)$;

(v) SDF (strongly distribution free) wrt $\mathcal{G}$ and $\Omega(H_0 \cup H_1)$ if $P\{A \mid H\} = P\{g(A) \mid H\}$ for all g in $\mathcal{G}$ and H in $\Omega(H_0 \cup H_1)$.

The rank sets are now defined as follows:

DEFINITION 2.5.

$$B(\pi, \pi') = \{(x, y): \pi(x_1) < \cdots < \pi(x_n); \pi'(y_1) < \cdots < \pi'(y_n)\},$$

the *B-rank set* generated by π and π'; and $B(\delta) = \cup B(\pi, \pi\delta)$, where (for fixed δ in $\mathcal{S}_n$) the union is over π in $\mathcal{S}_n$, *the* BNS-*rank set* generated by δ.

It is immediate that

(2.5) There are $(n!)^2$ B-rank sets each similar of size $(n!)^{-2}$; and $(n!)$ BNS-rank sets each similar of size $(n!)^{-1}$ wrt $\Omega(H_0)$. Further, each of these sets is invariant under $\mathcal{G}$; almost invariant wrt $\mathcal{G}$ and $\Omega(H_0 \cup H_1)$; and SDF wrt $\mathcal{G}$ and $\Omega(H_0 \cup H_1)$.

More generally:

(2.6) Each almost invariant set in SDF; and each SDF set is DF.

For the other two types of sets one needs the concept of Pitman statistics $T(T^*)$ defined in (F) below. The idea here is simply to use the B(NS)-Pitman function to select the $(n!)^2\alpha$ points for an α-Scheffé set.

DEFINITION 2.6. $\{T^*(v^*) = r\}$ is a *B-Pitman set* for $r = 1, 2, \cdots, (n!)^2$ and $\{T(v) = r\}$ is a BNS-*Pitman set* for $r = 1, 2, \cdots, (n!)$ if v^* and v are respectively, B-Pitman and BNS-Pitman functions and T and T^* are as defined below.

It can be proved that

(2.7) each B(BNS)-Pitman set is an α-Scheffé set for $\alpha = (n!)^{-2}((n!)^{-1})$, and that

(2.8) there exist B(BNS)-Pitman sets which are not B(NS)-rank sets and which are not invariant, SDF or almost invariant.

One can now introduce the statistics to be used.

(F) *Statistics.* A (nonrandomized) statistic here is as usual a real-valued measurable function W on R_{2n}.

DEFINITION 2.7. A statistic W is

(i) DF wrt $\Omega(H_0)$ if there exists a single distribution $M(W; \cdot)$ with the property that $P\{W(X, Y) \leqq t \mid H\} = M(W; t)$ for all H in $\Omega(H_0)$ and all t; and

(ii) SDF wrt $\mathcal{G}$ and $\Omega(H_0 \cup H_1)$ if for all t and H in $\Omega(H_0 \cup \mathrm{H}_1)$,

$$P\{W(X, Y) \leqq t \mid \hat{H}\} = P\{W(X, Y) \leqq t \mid H\}$$

whenever there exist g in $\mathcal{G}$ such that $\hat{H}g = H$, i.e. whenever H and $\hat{H}$ are elements of the same equivalence class of $\Omega(H_0 \cup H_1)$ under $\mathcal{G}$. [Note: For (ii) it is equivalent to say that the distribution of W is a function only of the maximal invariant in $\Omega(H_0 \cup H_1)$ under $\mathcal{G}$.]

One can readily demonstrate that

(2.9) W is DF(SDF) iff for each Borel set B in R_1, $W^{-1}(B)$ is DF(SDF).

For rank statistics one now defines

DEFINITION 2.8. W is a

(i) (NS) *rank statistic* if W is constant over the B(BNS)-rank sets;

(ii) B(BNS)-*rank statistic* if it is a (NS)-rank statistic and assumes $(n!)^2[(n!)]$ distinct values.

It follows then that

(2.10) Each rank statistic W is DF, SDF; invariant under $\mathcal{G}$; and that for each Borel set B, $W^{-1}(B)$ is almost invariant.

Finally one has the Pitman (or permutation) statistics

DEFINITION 2.9. (i) $T^*(v^*)$ is the *B-Pitman statistic* induced by B-Pitman function v^*, if for all z, $T^*(v^*(z)) = \sum \epsilon\{v^*(z) - v^*(\gamma(z))\}$, where the summation is over γ in $\mathcal{S}$ and ϵ is the function defined by $\epsilon(w) = 1$ or 0 according as $w \geqq 0$ or $w < 0$;

(ii) $T(v)$ is the BNS-*Pitman statistic* induced by BNS-Pitman function v, if for all $z = (x, y)$, $T(v(x, y)) = \sum \epsilon\{v(x, y) - v(x, \pi(y))\}$, where the summation is over π in $\mathcal{S}_n$;

(iii) $W^*[W]$ is a (NS)-*Pitman statistic* if it is a real-valued function of a B(NS)-Pitman statistic.

It is clear that

(2.11) $T^*(v^*)[T(v)]$ assumes only the values $r = 1, \cdots, (n!)^2[(n!)]$.

One should note here that in a NS-procedure it seems reasonable to use only NS statistics, but that there are other DF statistics available, e.g. the B-rank and B-Pitman statistics.

Finally, in this section one wishes to introduce a generalization of the ideas in [4] and (5) and several other authors.

(G) *Randomized statistics*. Randomized statistics L as defined below will not be statistics in the sense above. However, it will be seen in the sequel that test functions based on such statistics are DF statistics.

Let

(a) $\xi = (\xi_1, \cdots, \xi_n)$ and $\eta = (\eta_1, \cdots, \eta_n)$ be independent random samples both independent of the original data sample $Z = (X, Y)$;

(b) ξ_i and η_j have distributions F and G in $\Omega_2(R_1)$, respectively;

(c) M a desired distribution and f a statistic such that $P\{f(\xi, \eta) \leqq t \mid H\} = M(t)$ for all t;

(d) v^* a desired B-Pitman function; and

(e) $\bar{\gamma}(\cdot, \cdot) = \bar{\gamma}(\cdot, \cdot \mid v^*, z)$ such that $T^*(v^*(\bar{\gamma}(\xi, \eta))) = T^*(v^*(x, y))$, i.e. $\bar{\gamma}$ is a function of both v^* and the data point $z = (x, y)$, and as such is a random permutation.

DEFINITION 2.10. The *randomized statistic* induced by f, F, G and v^* is defined so that

$$L(x, y; \xi, \eta; v^*) = f(\bar{\gamma}(\xi, \eta)).$$

One notes here that the sets $\{T^*(v^*) = r\}$ constitute a partition of the sample space, and that the random permutation γ preserves the "partition information," i.e. the value of v^*, the function inducing the partition.

One now begins to construct the DF statistics, sets and tests.

3. Structure and invariance. In this section it will be shown that the similar test functions and sets, are respectively, the α-Lehmann-Stein functions and α-Scheffé sets; and that these can always be expressed in terms of B-Pitman functions. Further, the rank sets will be shown to be exactly those which are SDF and almost invariant.

THEOREM 3.1. (*Similar test functions*) (i) *φ is a size α test function similar* wrt $\Omega(H_0)$ iff *φ is an α-Lehmann-Stein function.*

(ii) *The following three conditions are equivalent:*

(a) *A is a set similar* wrt $\Omega(H_0)$ *and* $P(A) = \alpha$;

(b) $\varphi = I(A)$ *is a classical test of size* α; *and*

(c) *A is an α-Scheffé set.*

PROOF. (i) is a slight modification of a lemma of Lehmann (1959), p. 184. (ii) follows from (i) if one requires that φ be an indicator function.

For rank and Pitman sets one has

LEMMA 3.1. (i) *Each* B-*rank set and each* B-*Pitman set is similar* wrt $\Omega(H_0)$ *and has probability* $(n!)^{-2}$. (ii) *Each* BNS-*rank set and each* BNS-*Pitman set is similar* wrt $\Omega(H_0)$ *and has probability* $(n!)^{-1}$.

PROOF. (i) Each B-Pitman and each B-rank set contains exactly one point of a.a. orbits. Consequently, they are α-Scheffé sets for $\alpha = (n!)^{-2}$. (ii) The BNS-rank sets and BNS-Pitman sets contain exactly $(n!)$ points of a.a. orbits, and, hence, are α-Scheffé sets for $\alpha = (n!)^{-1}$.

From this lemma one gets the major structure theorem.

THEOREM 3.2. (*Nonrandomized similar statistics*) (i) *Each* (*nonrandomized*) DF *statistic has a discrete distribution with probabilities which are integral multiples of* $(n!)^{-2}$. *If, further, the statistic is* NS, *then the probabilities are multiples of* $(n!)^{-1}$.

(ii) W^* (*nonrandomized*) *is* DF wrt $\Omega(H_0)$, *iff* W^* *is equivalent to a function of a* B-*Pitman statistic,* i.e. *iff there exists a* B-*Pitman function* v^* *and a measurable function* U^* *such that* $W^* \equiv U^*[T^*(v^*)]$.

(iii) W(*nonrandomized*) *is* NS *and* DF wrt $\Omega(H_0)$ *iff* W *is equivalent to a* BNS-*Pitman statistic,* i.e. *iff there exists a* BNS-*Pitman function* v *and a measurable function* U *such that* $W = U[T(v)]$.

(iv) *For any preassigned discrete distribution* F, *with probabilities which are integer multiples of* $(n!)^{-2}$, *there exists a* DF *statistic with distribution* F.

PROOF. (i) $F(a) = P(W \leqq a)$ can assume only values $r(n!)^{-2}$ for $r = 0, 1, \cdots, (n!)^2$ since each set $(W \leqq a)$ is similar. If, further, W is BNS, then $F(a)$ can only assume the values $r(n!)^{-1}$ for $r = 0, 1, \cdots, n!$.

(ii) If v^* is a B-Pitman function, each set $\{T^*(v^*) = r\}$ is similar; and each set $\{U^*[T^*(v^*)] = r\}$ is also similar. Therefore W^* is DF. Conversely, if W^* is DF wrt $\Omega(H_0)$, then W^* has a discrete distribution; thus there exists $1 \leqq l \leqq (n!)^2$, numbers $a_1 < a_2 < \cdots < a_l$, and integers $0 = k(0) < k(1) < \cdots < k(l) = (n!)^2$ such that $P(W^* < a_1 \mid H_0) = 0$ and $P(W^* \leqq a_i \mid H_0) = P(W^* < a_{i+1} \mid H_0) = k(i)(n!)^2$. Consequently, for a.a. z, there exists an ordering $\{z_1, \cdots, z_s;\ s = (n!)^2\}$ of the points of $\mathcal{S}(z)$ such that $W^*(z_j) = a_i$ for $k(i-1) < j \leqq k(i)$. Define $v^*(z_j) = j$. Then $T^*[v^*(z_j)] = j$. Next define $U^*(j) = a_i$ for $k(i-1) < j \leqq k(i)$. Clearly v^* is a B-Pitman function (and a ranking function) and $W^* = U^*[T^*(v^*)]$.

(iii) The proof here is analogous to that of (ii).

(iv) If F assigns probabilities $p(1), \cdots, p(s)$ to the points $a(1), \cdots, a(s)$, where $\sum p(i) = 1$, then there exist integers $0 = k(0) < k(1) < \cdots < k(s) = (n!)^2$ such that $\sum_1^j p(i) = k(j)(n!)^{-2}$ for all j. Let v^* be an arbitrary B-Pitman function, and define $W = a(i)$ on each of the sets $\{T^*(v^*) = r\}$ for $k(i-1) < r \leqq k(i)$. Then W^* is a DF statistic with distribution F.

One has more distribution flexibility with randomized statistics.

THEOREM 3.3. (*Randomized similar statistics*) (i) *If φ is a test function based on a randomized statistic L* (i.e. $\varphi = 1$ *or* 0 *according as* $L \varepsilon A$ *or not for appropriate* A) *then φ is an α-Lehmann-Stein function and is, hence, similar* wrt $\Omega(H_0)$.

(ii) *For each preassigned distribution F, there exists a randomized statistic with F as its null hypothesis distribution.*

PROOF. (i) One recalls that $L(x, y, \xi, \eta; v^*) = f[\bar{\gamma}(\xi, \eta)]$, where $\bar{\gamma}(\cdot, \cdot) = \bar{\gamma}(\cdot, \cdot \mid v^*, z)$ is such that $T^*[v^*(z)] = T^*(v^*(\bar{\gamma}(\xi, \eta)))$. Therefore, if $\alpha = P\{f(\xi, \eta) \varepsilon A\}$, one has for a.a. orbits:

$$\begin{aligned}\textstyle\sum_\pi \varphi(\pi(z)) &= \textstyle\sum_\pi P\{\text{Reject } H_0 \mid \pi(z)\} = \sum_\pi P\{L(\pi(z), \xi, \eta; v^*) \varepsilon A\} \\ &= \textstyle\sum_\pi P\{f[\bar{\gamma}(\xi, \eta)] \varepsilon A \mid \pi(z)\} = \sum_\gamma P\{f[\gamma(\xi, \eta)] \varepsilon A\} \\ &= \textstyle\sum_\gamma P\{f(\xi, \eta) \varepsilon A\} = (n!)^2\alpha .\end{aligned}$$

(ii) Let the ξ sample have common distribution H which is strictly increasing and continuous; and let $L = F^{-1}H(\bar{\gamma}(\xi_1))$, where $\bar{\gamma}$ is the transformation in the definition of a randomized statistic; L, then, has the desired distribution F.

For the rank set result one needs the following two lemmas.

LEMMA 3.2. $\Omega(H_0 \cup H_1)$ *is* NS *complete.*

PROOF. $\Omega(H_0 \cup H_1)$ is the n-fold power class of $\Omega_2(R_2)$. From the main theorem of Bell, Blackwell and Breiman (1960) it is known that $\Omega_2(R_2)$ is symmetrically complete, i.e. if $\hat{h}(z_1, \cdots, z_n)$ is invariant under $\mathcal{S}_n$ and $\int \hat{h}\, dH^{(n)} = 0$ for all H in $\Omega_2(R_2)$, then $P\{\hat{h} \neq 0 \mid H^{(n)}\}^n = 0$ for all H in $\Omega_2(R_2)$. If one writes $z_i = (x_i, y_i)$ and $(z_1, \cdots, z_n) = (x_1, \cdots, x_n; y_1, \cdots, y_n)$, and $h(x_1, \cdots, x_n; y_1, \cdots, y_n) = \hat{h}(z_1, \cdots, z_n)$, one sees immediately that h is NS iff $\hat{h}$ is symmetric. Consequently, $\Omega(H_0 \cup H_1)$ is NS complete.

LEMMA 3.3. *If a* NS *similar set A is* SDF wrt $\Omega(H_0 \cup H_1)$ *and* $\mathcal{G}$, *then A is almost invariant* wrt $\mathcal{G}$ *and* $\Omega(H_0 \cup H_1)$.

PROOF. If $f, g \varepsilon \mathcal{G}_2$, and A is a NS similar set, then $P = P\{A \mid H\} =$

$P\{A \mid H(f, g)^{(n)}\} = P\{(f, g)^{(n)}(A) \mid H\}$ for all H in $\Omega(H_0 \cup H_1)$, and all $f, g \in \mathcal{G}_2$. Therefore, $\int \{I(A) - I[(f, g)^{(n)}(A)]\}\, dH = 0$ for all H in $\Omega(H_0 \cup H_1)$. But since $\Omega(H_0 \cup H_1)$ is NS complete, $P\{A \triangle (f, g)^{(n)}(A) \mid H\} = 0$ for all H in $\Omega(H_0 \cup H_1)$, and all $f, g \in \mathcal{G}_2$ where $A \triangle B = (A - B) \cup (B - A)$. Therefore, the conclusion follows.

The characterization result for rank statistics is now:

THEOREM 3.4. (*Rank Statistics*) (i) *W^* is a rank statistic iff it is equivalent to a function of the Pitman statistic of some ranking function*, i.e. *iff there exists measurable U^* and a ranking function v^* such that $W^* = U^*[T^*(v^*)]$.*

(ii) *The following conditions are equivalent:*

(a) *W is a* NS *rank statistic,*

(b) *W is* NS *and* SDF wrt $\Omega(H_0 \cup H_1)$ *and* $\mathcal{G}$, *and,*

(c) *W is equivalent to a* NS-*Pitman statistic of some* NS *ranking function*, i.e. *there exists measurable U and a* NS-*ranking function v such that $W = U[T(v)]$.*

PROOF. (i) From Theorem 3.2 it follows that each rank statistic is equivalent to some Pitman statistic, since a rank statistic is a DF statistic. Each rank statistic W can be represented as a function U of a rank statistic W_1^* which assumes different values on each of the $(n!)^2$ B-rank sets. This latter statistic is DF and, hence, can be written $W_1^* = U_1^*[T^*(v^*)]$, and in this case, v^* is necessarily a ranking function. Consequently, $W = U[U_1^*[T^*(v^*)]]$. Conversely, if v^* is a ranking function, each function of $T^*(v^*)$ is constant over the B-rank sets, and is, therefore, a rank statistic.

(ii) The proof of the equivalence of (a) and (c) parallels the proof of (i) above.

Since the B-rank sets $B(\pi, \pi')$ are invariant under $\mathcal{G}$, it follows that the BNS-rank sets are also invariant and, hence, are SDF wrt $\Omega(H_0 \cup H_1)$ and $\mathcal{G}$. Hence any NS rank statistic is SDF wrt $\Omega(H_0 \cup H_1)$ and $\mathcal{G}$, i.e. (a) implies (b). To show that (b) *implies* (a), it is sufficient to prove that if A is NS and is SDF, its intersection with an arbitrary B-rank ($B(\pi, \pi')$) set is equivalent either to the null set or to the B-rank set.

$B(\pi, \pi')$, being an invariant set, is almost invariant wrt $\mathcal{G}$ and $\Omega(H_0 \cup H_1)$. By Lemma 3.3, A is almost invariant. But the class of almost invariant sets is closed under intersections, and is a sub-family of the sets similar wrt $\Omega(H_0)$. $A \cap B(\pi, \pi')$ is, therefore, similar wrt $\Omega(H_0)$ and, since, also it is a subset of $B(\pi, \pi')$, it is an α-Scheffé-set with $\alpha = (n!)^{-2}$ or 0. In the first case, it is equivalent to $B(\pi, \pi')$ and in the second case it is equivalent to the empty set.

Now that the families of DF statistics are known one wishes to select those which are optimal in some sense.

4. Optimality for parametric alternatives. From [21] and [20], p. 185, one has:

LEMMA 4.1. (Lehmann and Stein). *In the class of all* DF *tests, the* MP (*most powerful*) *level α test of H_0 against H_1 is of the form:*

$$\begin{aligned} \varphi &= 1 \quad \text{if} \quad T[v(z)] > k \\ &= \lambda \quad \text{if} \quad T[v(z)] = k \\ &= 0 \quad \text{otherwise} \end{aligned} \tag{4.1}$$

where v is the density of Z under H_1 and $T[v(z)] = \sum_\pi \epsilon\{v(x, y) - v(\pi(x), y)\}$.

[Note: Although v is NS, v need not be a BNS-Pitman function (Example 4.1); however, there exists a BNS-Pitman function v' which generates the same test.]

EXAMPLE 4.1. Consider the model (Konijn (1956)) in which X and Y can be written $X = U + \theta V$ and $Y = \theta U + V$, where $0 \leqq \theta \leqq 1$, U and V are independent with continuous distributions F_0 and G_0 respectively. The MP level α DF test of $\theta = 0$ vs. $\theta > 0$ rejects for large values of $T[v(z)]$ with

$$v(z) = \sum [\ln f_0((X_i - \theta y_i)/(1 - \theta^2)) + \ln g_0((y_i - \theta x_i)/(1 - \theta^2))]$$

when F_0 and G_0 have the positive densities f_0 and g_0 respectively. If $f_0(x) = g_0(x) = \exp(-x)$ when $x \geqq 0$, $= 0$ otherwise (i.e. both are standard exponential), then $v(z)$ of the MP test can be written $v(z) = \epsilon[\min_i (x_i - \theta y_i)]\epsilon[\min_i (y_i - \theta x_i)]$. If f_0 and g_0 are normal densities, then $v(z) = \sum x_i y_i$.

[Note: In the exponential case above v assumes only the two values 0 and 1. This is the worst possible case under random sampling since v can only be constant under the null hypothesis. One sees that $\prod_1^n F(x_i, y_i) = \prod_1^n F(x_i, \delta(y_i))$ for all δ in $\mathcal{S}_n$ and all x_i and y_i iff $F(x_i, y_i) = F(x_i, \infty)F(\infty, y_i)$ for all x_i and y_i. This means that the MP DF test for any given parametric alternative has power greater than α.]

The MP tests in the above example typically depend on θ and on F_0 and G_0. In order to obtain tests that do not depend on θ and still are optimal in some sense, one introduces the concept of locally MP tests (e.g. [25] and [20]).

To this end one needs the following notation:

(a) A, an interval containing 0;

(b) $\{Q(\theta; \cdot, \cdot): \theta \varepsilon A\}$, a class of absolutely continuous bivariate distributions with

(i) $Q(0; x, y) = Q_1(x)Q_2(y)$ for all x and y; and

(ii) regularity conditions given below in terms of the power function β_φ;

(c) $h(\theta; z) = \prod_i q(\theta; x_i, y_i)$, where q is the density of Q;

(d) $\beta_\varphi(\theta)$, the power of the test φ against alternative F_θ; and

(e) $\beta_\varphi^{(r)}$, the rth derivative wrt θ of $\beta_\varphi(\theta)$.

It is also worthwhile to mention that the DF property refers to $\Omega(H_0)$, i.e., the distribution of the test statistic is invariant over $\Omega(H_0)$. In the sequel, one seeks an additional property—that the DF test be "good" against the entire family $\{Q(\theta; \cdot, \cdot): \theta \varepsilon A\}$ of alternatives.

DEFINITION 4.1. A level α test φ_0 is *locally* MP for testing $\theta = 0$ vs. $\theta > 0$ if, given any other level α test φ_1 there exists $\Delta(\varphi_1)$ such that

$$\beta_{\varphi_0}(\theta) \geqq \beta_{\varphi_1}(\theta) \quad \text{for all } \theta \text{ with } 0 < \theta < \Delta(\varphi_1). \tag{4.2}$$

THEOREM 4.1. *If there exist $a > 0$ such that for all $\theta \varepsilon (-a, a)$, all $z \varepsilon R_{2n}$, and all level α DF tests φ*

(i) *$h^{(r)}(\theta; z) = (\delta^r/\delta\theta^r)h(\theta; z)$ exists and is continuous, and*

(ii) *$\beta_\varphi^{(r)}(\theta) = (\delta^r/\delta\theta^r)\beta_\varphi(\theta)$ exists; is continuous, and can be obtained by differentiating inside the integral sign in (4.5); and if*

(iii) $r \geqq 1$ *is the smallest integer for which* $h^{(r)}(0; z)$ *is not invariant* wrt $\mathcal{S}_n \times \mathcal{S}_n$; *then the locally* MP DF *level* α *test of* $\theta = 0$ *against* $\theta > 0$ *is of the form* (4.1) *with*

$$v(z) = h^{(r)}(0; z). \tag{4.3}$$

PROOF. It is known (e.g. [21] and [20, p. 342]) that it is sufficient to maximize $\beta_\varphi^{(r)}(0)$ subject to

$$\sum_\gamma \varphi[\gamma(z)] = (n)^2\alpha. \tag{4.4}$$

Following Lehmann (1959), p. 185, one writes

$$\beta_\varphi(\theta) = \int [\sum_\gamma \varphi[\gamma(t)]h(\theta; \gamma(t))/\sum_\gamma h(\theta; \gamma(t))]\, dP^T(t), \tag{4.5}$$

where T is the order statistic vector $(X(1), \cdots, X(n); Y(1), \cdots, Y(n))$. Hence, in order to maximize $\beta_\varphi^{(r)}(0)$ subject to (4.4), it is sufficient to maximize the rth derivative of the integrand in (4.5) subject to (4.4). Upon computing this derivative and applying the Neyman-Pearson lemma, the result follows.

"r" of Theorem 4.1 is equal to one or two for most models, and $v(z)$ in (4.3) can be written in the equivalent form

$$v(z) = \sum (\delta^r/\delta\theta^r) \ln q(\theta; x_i, y_i) \mid \theta = 0 \tag{4.6}$$

when the logarithms used are finite.

EXAMPLE 4.2. For the model of Example 4.1, $r = 1$ and the locally MP DF test rejects for large values of $T[v(z)]$ with $v(z) = -\sum \{[f_0'(x_i)y_i/f_0(x_i)] + [g_0'(y_i)x_i/g_0(y_i)]\}$. If $F_0(x) = G_0(x) = 1/[1 + \exp(-x)]$ (i.e. both are logistic), then $v(z)$ of the locally MP test may be written $\sum [F_0(x_i)y_i + F_0(y_i)x_i]$.

EXAMPLE 4.3. Consider the model (Bhuchongkul (1964)) in which $X = (1 - \theta)U + \theta W$ and $Y = (1 - \theta)V + \theta W$ where $0 \leqq \theta \leqq 1$, U, V and W are independent with continuous distributions F_0, G_0 and F_1 respectively. For this model, $r = 2$, and $v(z)$ of Theorem 4.1 is equivalent to $\sum f_0'(x_i)g_0'(y_i)/f_0(x_i)g_0(y_i)$. If F_0 and G_0 are normal, this becomes $\sum x_iy_i$, and if F_0 and G_0 are logistic, it can be written $\sum F_0(x_i)F_0(y_i)$.

EXAMPLE 4.4. For the model of Jogdeo (1964), $Y = \theta X + \sigma W$, where $\theta \geqq 0$ and X and W are independent and have the continuous distributions F and F_1 respectively. When F_1 has the positive density f_1, the MP DF test is of the form (4.1) with $v(z) = \sum \ln f_n[(y_i - \theta x_i)/\sigma]$. When f_1 is exponential, then $v(z) = \epsilon[\min_i (y_i - \theta x_i)]$, and when f_1 is normal, $v(z) = \sum x_iy_i$. The locally MP DF test is of the form (4.1) with $v(z) = -\sum f_1'(y_i)x_i/f_1(y_i)$ which becomes $\sum F_1(y_i)x_i$ when F_1 is logistic.

The MP rank (SDF) test for H_0 vs. H_1 can be derived by the usual (e.g. [20], p. 237) arguments and it is found that if $r_i(s_i)$ denotes the rank of $x_i(y_i)$ in the $x(y)$ sample, $\xi(i)$ ($\eta(i)$) is the ith order statistic in a random sample of size n from a population with density $h_x(h_y)$, and the ξ, η, and z samples are mutually independent, then

LEMMA 4.2. *If* h_xh_y *is positive whenever* q *is, then the* MP *rank test of* H_0 vs. H_1 *is of the form* (4.1) *with*

$$v(z) = E\{\prod_{i=1}^n q[\xi(r_i), \eta(s_i)]/h_x[\xi(r_i)]h_y[\eta(s_i)]\}. \tag{4.7}$$

Locally MP rank tests have been given by Bhuchongkul for certain models. Using standard (e.g. [10], [5]), arguments, one can show that

LEMMA 4.3. *If the condition of Lemma* 4.2 *holds and if the conditions of Theorem* 4.1 *hold with* (4.5) *replaced by* (4.7), *then the locally* MP *rank test of* $\theta = 0$ vs. $\theta > 0$ *is of the form* (4.1) *with*

$$v(z) = E\{(\partial^r/\partial\theta^r)\prod_{i=1}^{n} q[\theta;\, \xi(r_i),\, \eta(s_i)]_{\theta=0}\}. \tag{4.8}$$

Note that $h_x h_y = q(0;\, \cdot,\, \cdot)$ and that (4.8) can be written

$$v(z) = E\{\sum_{i=1}^{n} (\partial^r/\partial\theta^r) \ln q[\theta;\, \xi(r_i),\, \eta(s_i)]_{\theta=0}\} \tag{4.9}$$

when the logarithms involved are finite.

Because of the paralellism between (4.9) and (4.6), the locally MP rank statistics for the Konijn, Bhuchongkul and Jogdeo models can be obtained from Examples 4.2, 4.3 and 4.4.

5. Randomized statistics.

Although randomized statistics can be defined from arbitrary Pitman (DF) statistics, they are more practical when defined from rank (SDF) statistics. The expectations that appear in the MP rank statistic (4.7) and locally MP rank statistic (4.9) are often not available, and one may want to approximate them by what is left when the expectation signs are removed. This leads to randomized statistics of the form

$$nK_n = \sum_i \xi(r_i)\eta(s_i), \tag{5.1}$$

$$nK_n' = \sum_i (\xi(r_i) - \mu_1)(\eta(s_i) - \mu_2), \tag{5.2}$$

$$n\hat{K}_n = \sum_i (\xi(r_i) - \bar{\xi})(\eta(s_i) - \bar{\eta}) \tag{5.3}$$

where $\xi(i)(\eta(i))$ is the ith order statistic in a random sample of size n form a continuous distribution $F(G)$; the ξ and η samples are independent and independent of the (X, Y) sample; $\mu_1(\mu_2)$ is the mean of $F(G)$; and $\bar{\xi}(\bar{\eta})$ is the sample mean of the $\xi(\eta)$ sample.

One advantage of the first two statistics above is that they are the sums of independent identically distributed random variables and their distributions are therefore easily obtainable (see [4]). Randomized statistics similar to the above have been considered by Ehrenberg (1951) and Durbin (1961). Some of their properties are developed in [4].

The purpose of this section is to show that K_n' and $\hat{K}_n$ are in some sense asymptotically equivalent to the locally MP rank statistic,

$$nT_n = \sum_i E[\xi(r_i)]E[\eta(s_i)], \tag{5.4}$$

while K_n and T_n are asymptotically equivalent iff $\mu_1 = \mu_2 = 0$.

THEOREM 5.1. (i) *If F and G have first moments μ_1 and μ_2, then*

$$E(K_n \mid H_0 \cup H_1) = E(T_n \mid H_0 \cup H_1) = E(K_n' \mid H_0 \cup H_1) - \mu_1\mu_2$$
$$= E(\hat{K}_n \mid H_0 \cup H_1) - \mu_1\mu_2 .$$

(ii) *If F and G have variances σ_1^2 and σ_2^2, then $E(n(K_n' - T_n + \mu_1\mu_2)^2 \mid H_0)$ and $E(n(\hat{K}_n - T_n + \mu_1\mu_2)^2 \mid H_0)$ tend to zero as $n \to \infty$, while $n^{\frac{1}{2}}(K_n - T_n)$ tends in law to a normal variable with mean zero and variance $\mu_1^2\sigma_2^2 + \mu_2^2\sigma_1^2$.*

PROOF. Define $z_{ij} = 1$ if there exist $1 \leqq k \leqq n$ such that $(r_k, s_k) = (i, j)$, $z_{ij} = 0$ otherwise, then

$$(5.5) \quad \hat{K}_n = n^{-1}\textstyle\sum_{i,j} \xi(i)\eta(j)[z_{ij} - n^{-1}] = K_n - \bar{\xi}\bar{\eta} = K_n' - (\bar{\xi} - \mu_1)(\bar{\eta} - \mu_2),$$

$$T_n = n^{-1}\textstyle\sum_{i,j}E[\xi(i)]E[\eta(j)][z_{ij} - n^{-1}] + \mu_1\mu_2.$$

By definition, the ξ's and η's are independent of the z's, thus $E\hat{K}_n = EK_n' = EK_n - \mu_1\mu_2 = ET_n - \mu_1\mu_2$, and (i) follows.

Next one shows that $E(D_n^2 \mid H_0) \to 0$ where $D_n^2 = n(\hat{K}_n - T_n + \mu_1\mu_2)$ and $E(D_n^2 \mid H_0) = n^{-1}\sum_{ijkm} \operatorname{Cov}[\xi(i)\eta(j), \xi(k)\eta(m)] \operatorname{Cov}(z_{ij}, z_{km})$.

From the elementary inequality $|\operatorname{Cov}(X, Y)| \leqq \sigma_x\sigma_y \leqq \frac{1}{2}(\sigma_x^2 + \sigma_y^2)$ and the fact that $(n - 1)^{-1} \operatorname{Var}(z_{ij}) = -\operatorname{Cov}(z_{ij}, z_{im}) = -\operatorname{Cov}(z_{ij}, z_{kj}) = (n - 1)\operatorname{Cov}(z_{ij}, z_{km}) = n^{-2}$ when $i \neq k$ and $j \neq m$, one obtains $E(D_n^2 \mid H_0) \leqq (2.5)\, n^{-2}\sum_{ij} \operatorname{Var}[\xi(i)\eta(j)]$. Since

$$\textstyle\sum_{ij} \operatorname{Var}[\xi(i)\eta(j)] = \sum_{ij} E(\sum_i \xi_i^2)E(\sum_j \eta_j^2) - \sum_i [E\xi(i)]^2 \sum_j [E\eta(j)]^2$$
$$= n^2E\xi^2E\eta^2 - \textstyle\sum_i[E\xi(i)]^2\sum_j[E\eta(j)]^2,$$

it now follows from the results of Hoeffding (1953) that $E(D_n^2 \mid H_0)$ tends to zero. The remainder of (ii) follows from (5.5) and the equation

$$K_n - T_n = (K_n' - T_n + \mu_1\mu_2) + \mu_1(\bar{\eta} - \mu_2) + \mu_2(\bar{\xi} - \mu_1).$$

Theorem 5.1 implies that $T_n - \mu_1\mu_2$ and K_n' have the same asymptotic null-distribution. Since K_n' is a sum of independent identically distributed random variables, one has the following result for the nonrandomized rank statistics T_n.

COROLLARY 5.1. *If F and G have second moments then $n^{\frac{1}{2}}[T_n - \mu_1\mu_2]/\sigma_1\sigma_2$ has asymptotically a standard normal distribution under H_0.*

Applying the theory of contiguity developed by Le Cam (1960) and Hájek (1962), one may use Theorem 5.1 to show that T_n and K_n' are asymptotically equivalent also under local alternatives. Let $H_1^{(n)}$ denote an alternative $Q(\theta_n; x, y)$ $(\theta_n \to 0)$ for which the contiguity conditions are satisfied (see [16], [17] and [19]), then

COROLLARY 5.2. *$P(n^{\frac{1}{2}}|K_n' - T_n + \mu_1\mu_2| \geqq \epsilon \mid H_1^{(n)}) \to 0$ as $n \to \infty$ for all $\epsilon \geqq 0$ provided F and G have second moments. In particular, K_n' and T_n have asymptotic relative* (Pitman) *efficiency one for $Q(\theta_n; x, y)$.*

Note that (5.5) implies that $n^{\frac{1}{2}}(K_n' - \hat{K}_n)$ tends to zero for any alternative provided that F and G have second moments.

6. Asymptotically minimax tests. In Section 4, locally optimal DF statistics which do not depend on the parameter θ are obtained. However, these statistics depend on the underlying distributions (e.g. F_0 and G_0 in Example 4.3). The purpose of this section is to use the ideas of [13] to obtain optimal statistics which do not have this dependence. This is achieved by asymptotically maximizing the minimum power over classes of distributions, i.e. by obtaining asymptotic minimax solutions.

Let $\mathfrak{F}(\sigma)$ denote the class of univariate distributions F with a twice differentiable density f and with a variance $\sigma^2(F) \leqq \sigma^2$, let $\mathfrak{F}_0$ be the class of all continuous distributions with a finite variance, and let $\beta_\varphi(F_0, G_0, F_1; \theta)$ denote the power of each test φ for the Bhuchongkul model of Example 4.3, then the results of Chernoff and Savage (1958), Bhuchongkul (1964), and [13] easily yield

Theorem 6.1. *The level α normal scores test φ_1 is asymptotically minimax in the sense that*

$$(6.1) \quad \inf [\liminf_n \beta_{\varphi_1}(F_0, G_0\ F_1; \theta_n)] = \sup_\varphi \{\inf [\limsup_n \beta_\varphi(F_0, G_0, F_1; \theta_n)]\}$$

where the infima are over F_0, G_0 in $\mathfrak{F}(\sigma)$ and F_1 in $\mathfrak{F}_0$; the supremum is over the class $\mathfrak{A}^*$ of all level α tests; and the equality holds for all sequences θ_n which satisfy $\lim \theta_n n^{\frac{1}{4}} = c$ *for some* $0 \leqq c \leqq \infty$.

Proof. Note that the results of Chernoff and Savage (1958), Bhuchongkul (1964), and Fatou's lemma imply that

$$(6.2) \quad \inf [\liminf_n \beta_{\varphi_1}(F_0, G_0, F_1; \theta_n)] = \lim_n \beta_{\varphi_1}(\Phi, \Phi, \Phi_b; \theta_n)$$

where Φ_b denotes the normal distribution with mean zero and variance b^2, $\Phi = \Phi_1$ (without loss of generality, one sets $\sigma^2 = 1$), and the right hand side is independent of b. It is moreover known (e.g. [7]) that when $(F_0, G_0, F_1) = (\Phi, \Phi, \Phi_b)$, then φ_1 has the same asymptotic power as the test φ_0 that rejects for large values of $\sum x_i y_i$. One can easily show that for this model with $b^2 = (2 - \theta)/\theta$, φ_0 is locally MP and asymptotically efficient for testing $\theta = 0$ vs. $\theta > 0$. From these last two statements, one has

$$(6.3) \quad \inf_{b>0} [\limsup_n \beta_\varphi(\Phi, \Phi, \Phi_b; \theta_n)] \leqq \lim_n \beta_{\varphi_1}(\Phi, \Phi, \Phi_b; \theta_n)$$

for all tests that have asymptotically level α when $(F_0, G_0, F_1) = (\Phi, \Phi, \Phi_b)$.

Since Φ is in $\mathfrak{F}(1)$ and Φ_b is in $\mathfrak{F}_0$, then

$$(6.4) \quad \inf [\limsup_n \beta_\varphi(F_0, G_0, F_1; \theta_n)] \leqq \limsup_n \beta_\varphi(\Phi, \Phi, \Phi_b; \theta_n)$$

for all $b > 0$. Now (6.2), (6.3) and (6.4) yield (6.1).

$\mathfrak{A}$ does not include the classical correlation coefficient. However, (6.3) above shows that the result holds for the class $\mathfrak{A}'$ of all tests that have asymptotically levels α when X and Y are normally distributed with zero means and identical variances. $\mathfrak{A}'$ includes the correlation coefficient and contains $\mathfrak{A}$.

Corollary 6.1. *Theorem 6.1 holds if $\mathfrak{A}$ is replaced by $\mathfrak{A}'$.*

Furthermore, from (6.2) one can easily compute the following result in which Φ denotes the standard normal distribution and $k_\alpha = \Phi^{-1}(\alpha)$.

Corollary 6.2. *The quantities (6.1) of Theorem 6.1 equal $\Phi(k_\alpha + c^2/\sigma^2)$.*

Next one considers maximizing the minimum power over the class of alternatives $\Omega(a; d)$ defined below. This class is a natural extension of the Birnbaum alternatives [6], [8], [9], [11] in the one and two-sample cases to the bivariate independence case.

For convenience one introduces the following notation:

(a) $Y(x)$, the random variable with the conditional distribution of Y given $X = x$;

(b) $F(y \mid X = x) = P(Y(x) \leqq y)$;

(c) $\delta(Z, W) = \sup |F_Z(t) - F_W(t)|$, where the supremum is taken over all real t and F_Z and F_W are the distributions of Z and W, respectively, the Kolmogorov distance between random variables Z and W.

DEFINITION 6.1. For $0 < a, d < 1$, $\Omega(a; d)$ is the class of continuous bivariate alternatives H with the following properties:

(i) the distributions of the $Y(x)$'s are stochastically non-decreasing in x, i.e. $F(y \mid X = x) \geqq F(y \mid X = z)$ for all y whenever $x \leqq z$; and

(ii) $\delta(Y(x), Y(b)) \geqq d$ for all $x > b = F_X^{-1}(a)$.

$\mathfrak{J}$ will denote the class of Bhuchongkul statistics of the form $T(J_n, L_n) = n^{-1} \sum J_n(r_i/(n+1))L_n(s_i/(n+1))$, where (i) J_n and L_n converges to the functions J and L on $(0, 1)$; and (ii) J, L, J_n and L_n satisfies the regularity conditions of Theorem 1 of Bhuchongkul (1964).

DEFINITION 6.2. $\bar{\mathfrak{J}}$ (the *closure* of $\mathfrak{J}$) is the union of $\mathfrak{J}$ with the class of statistics that are of the form $T(J, L) = n^{-1} \sum J(r_i/(n+1))L(s_i/(n+1))$ and for which there exists a sequence $\{T(J^k, L^k)\}$ in $\mathfrak{J}$ such that J^k and L^k converges pointwise on $[0, 1]$ to J and L as $k \to 0$.

$\beta_T(H)$ will denote the power of the level α test with rejection region of the form $(T > c)$. The asymptotic minimum power of T at a is

$$(6.5) \qquad \beta_T(AM; a) = \lim_n (\inf [\beta_T(H) : H \,\varepsilon\, \Omega(a; d_n)])$$

whenever this limit exists; when it does not exist (or is not known to exist), one uses

$$(6.6) \qquad \hat{\beta}_T(AM; a) = \lim \sup_n (\inf [\beta_T(H) : H \,\varepsilon\, \Omega(s; d_n)]).$$

From considerations of symmetry, it is reasonable to restrict attention to the case where $a = \frac{1}{2}$ and J_n is symmetric in the sense that $J_n(u) = -J_n(1-u) + c_n$ for some constant c_n. One sets $\beta_T(AM) = \beta_T(AM; \frac{1}{2})$ and lets $\mathfrak{J}(S)(\bar{\mathfrak{J}}(S))$ denote the class of $T(J_n, L_n)$ in $\mathfrak{J}(\bar{\mathfrak{J}})$ for which J_n is symmetric.

Let $s_1, \cdots, s_n$ denote the y-ranks after the original samples have been permuted to make the x's ordered.

If $W = 2n^{-1}(n+1)^{-1} \sum_{\frac{1}{2}n+1}^{n} s_i - \frac{1}{2}$ when n is even and $W = 2n^{-1}(n+1)^{-1} \sum_{\frac{1}{2}(n+1)}^{n} s_i - \frac{1}{2} s_{\frac{1}{2}(n+1)} - \frac{1}{2}$ when n is odd; then the following result is proved in Section 7.

THEOREM 6.2. (i) *W is in $\bar{\mathfrak{J}}(S)$; (ii) $\beta_W(AM)$ exists and equals $\Phi(k_\alpha + 3^{\frac{1}{2}}c^2/2)$ iff $\lim_n (n^{\frac{1}{2}} d_n) = c$ for some $0 \leqq c \leqq \infty$; (iii) whenever $\lim_n (n^{\frac{1}{2}} d_n) = c$ for some $0 \leqq c \leqq \infty$; then W is asymptotically minimax over $\mathfrak{J}(S)$ and $\Omega(\frac{1}{2}; d_n)$ in the sense that*

$$(6.7) \qquad \beta_W(AM) = \sup [\hat{\beta}_T(AM) : T \,\varepsilon\, \mathfrak{J}(S)];$$

(iv) *if V is in $\mathfrak{J}(S)$ and $\lim_n (n^{\frac{1}{2}} d_n) = c$ for some $0 \leqq c \leqq \infty$, then $\beta_W(AM) \geqq \hat{\beta}_V(AM)$.*

Let $W_1 = \sum_{\frac{1}{2}n+1}^{n} s_i$ when n is even and $W_1 = \sum_{\frac{1}{2}(n+1)}^{n} s_i$ when n is odd. Then there are constants a_n and b_n such that $n^{\frac{1}{2}}(W - a_n W_1 - b_n) \to 0$ as $n \to \infty$, i.e. W_1 is asymptotically equivalent to W. The advantage of W_1 is given in the following observation.

LEMMA 6.1. *When n is even (odd), the null-distribution of W_1 is the same as the null-distribution of the two-sample Wilcoxon statistic based on samples of size $\frac{1}{2}n$ and $\frac{1}{2}n[\frac{1}{2}(n+1)$ and $\frac{1}{2}(n+1) - 1]$.*

7. Monotone tests and minimum power. Lower bounds for the power of distribution-free tests have been found by Birnbaum (1953), Chapman (1958), Birnbaum and Tang (1964), Bell, Moser and Thompson (1966), and others for the one-sample and two-sample problem. This section gives analogous lower bounds for the power of rank tests of independence.

$\Omega(a; d)$ is as in Section 6; the univariate distribution $B_1(r, d; t)$ is the Birnbaum (1953) alternative which coincides with the standard uniform distribution for t outside the interval $(r, r + d)$, and equals r for t in $(r, r + d)$. $B_1(r, d, e; t)(e > 0)$ is a continuous distribution on [0, 1] such that for each $0 < e_1 < e_2$, $B_1(r, d, e_2; t) < B_1(r, d, e_1; t) < B_1(r, d; t)$ and $B_1(r, d, e; t)$ converges to $B_1(r, d; t)$ as e tends to 0^+.

$B(r, d, e; x, y)$ is the bivariate distribution for which (i) $Y(x)$ is standard uniform for $x \leqq a$; (ii) $Y(x)$ has the distribution $B_1(r, d, e; t)$ for $x > a$; and (iii) the x-marginal is uniform. Note that $B(r, d, e; \cdot, \cdot)$ is in $\Omega(a; d)$.

Only non-sequential tests will be considered, thus one may assume that the data have been permuted so that $X_1 < X_2 < \cdots < X_n$.

DEFINITION 7.1. A *monotone* test function φ is such that $\varphi(x; y) \leqq \varphi(x; y')$ for all y, y' such that $y_i' < y_j'$ whenever $i < j$ and $y_i < y_j$.

Lehmann (1966) has shown that monotone NS tests are unbiased for alternatives such that $Y(x)$ is stochastically increasing. The following two lemmas are modifications of his arguments.

LEMMA 7.1. *If H is in $\Omega(a; d)$ and φ is a non-sequential monotone rank test, then there exists $0 \leqq r \leqq 1 - d$ and $e > 0$ such that $\beta_\varphi(B(r, d, e; \cdot, \cdot)) \leqq \beta_\varphi(H)$.*

PROOF. Let F_X, F_Y, K_x and K_b be the distributions of X_i, Y_i, $Y(x)$ and $Y(b)$, respectively; $\hat{H}(\cdot, \cdot) = H(F_X^{-1}(\cdot), K_b^{-1}(\cdot))$; $U_i = F_X(X_i)$, $V_i = K_b(Y_i)$; and $V(u)$ be analogous to $Y(x)$. Then, since φ is a rank test,

$$\begin{aligned}\beta_\varphi(H) &= E[\varphi(X; \mathrm{Y}) \mid H^{(n)}] \\ &= E\{\varphi[F_X(X_1), \cdots, F_X(X_n); K_b(Y_1), \cdots, K_b(Y_n)] \mid H^{(n)}\} \\ &= E[\varphi(U; V) \mid \hat{H}^{(n)}].\end{aligned}$$

Further, it is clear that the transformation of (X_i, Y_i) into (U_i, V_i) preserves the Kolmogorov distance δ in the sense that $\delta(Y(x), Y(b)) = \delta(V(u), V(a))$ for $b = F_X^{-1}(a)$. To see this, note that the distribution of $V(u)$ equals $K_x K_b^{-1}$ for $x = F_X^{-1}(u)$. Also note that $V(a)$ has a standard uniform distribution.

It is now convenient to let $U(r, d, e)$ denote a random variable whose distribution is $B_1(r, d, e; \cdot)$, then the following *stochastic* inequalities hold for some $0 \leqq r \leqq 1 - d$, $e > 0$, $V(u_1) \dot{\leqq} \cdots \dot{\leqq} V(u_k) \dot{\leqq} U_s \dot{\leqq} U(r, d, e) \dot{\leqq} V(u_{k+1}) \dot{\leqq} \cdots \dot{\leqq} V(u_n)$ where $k = \sum \epsilon(a - u_i)$ is the number of u's not greater than a, and U_s is a standard uniform variable. This follows by considering the distribution of $V(u)$ as given above and the results of [6] and [11]. From Lehmann (1959), p. 73, one finds the existence of non-decreasing functions f_i and g such that (i) $f_1(t) \leqq \cdots \leqq f_k(t) \leqq t \leqq g(t) \leqq f_{k+1}(t) \leqq \cdots \leqq f_n(t)$; (ii) the distribution of $f_i(U_s)$ equals the distribution of $V(u_i)(i = 1, \cdots, n)$; and (iii) the distribution of $g(U_s)$ equals the distribution of $U(r, d, e)$. Suppose that φ satisfies

$$(7.1)\quad \varphi(u; w_1, \cdots, w_k, g(w_{k+1}), \cdots, g(w_n)) \leqq \varphi(u; f_1(w_1), \cdots, f_n(w_n))$$

for all u such that $0 < u_1 < u_2 < \cdots < u_n < 1$, all w such that $0 < w_i < 1$, and each $k = 0, \cdots, n$; then

$$(7.2)\quad \varphi(U; W_1, \cdots, W_k, g(W_{k+1}), \cdots, g(W_n)) \leqq \varphi(U; f_1(W_1), \cdots, f_n(W_n))$$

where $W = (W_1, \cdots, W_n)$ and $U = (U_1, \cdots, U_n)$ are samples (U is ordered) from the standard uniform distribution. Upon writing $\beta_\varphi(H) = E_H\varphi = \sum_{i=0}^n E_H(\varphi \mid k = i) . P_H(k = i)$ and taking expectations with respect to W and U in (7.2), one finds

$$(7.3)\quad \beta_\varphi(B(r, d, e; \cdot, \cdot) \leqq \beta_\varphi(H).$$

It has now been shown that (7.3) holds for all tests satisfying (7.1). The fact that it holds for all monotone tests is immediate from the following result.

LEMMA 7.2. *The class of all tests satisfying* (7.1) *contains the class of monotone tests.*

PROOF. Let φ be a monotone test, then $\varphi(u; y) \leqq \varphi(u; y')$ for all (y, y') such that $y_i' < y_j'$ whenever $i < j$ and $y_i < y_j$. However, $w_i < w_j$ implies $f_i(w_i) < f_i(w_j) \leqq f_j(w_j)$ when $i < j \leqq k$; $w_k < g(w_j)$ implies $f_k(w_k) \leqq w_k < g(w_j) \leqq f_j(w_j)$ when $j > k$; and $g(w_i) < g(w_j)$ implies $w_i < w_j$ which yields $f_i(w_i) < f_i(w_j) \leqq f_j(w_j)$ when $k < i < j \leqq n$, thus in particular, φ satisfies (7.1).

From Lemma 7.1, one obtains

THEOREM 7.1. *If φ is a non-sequential monotone rank test, then*

$$\begin{aligned}\inf[\beta_\varphi(H): H \,\varepsilon\, \Omega(a; d)] &= \inf[\beta_\varphi(B(r, d, e; \cdot, \cdot)): 0 \leqq r \leqq 1 - d, e > 0] \\ &= \inf[\lim_{e\to 0+} \beta_\varphi(B(r, d, e; \cdot, \cdot)): 0 \leqq r \leqq 1 - d].\end{aligned}$$

One can now compute the asymptotic minimum power $\beta_R(AM; a)$ (see (6.5)) of the level α rank correlation coefficient test which rejects for large values of $R = n^{-1}(n + 1)^{-2} \sum r_i s_i$. Let $k_\alpha = \Phi^{-1}(\alpha)$, where Φ is the standard normal distribution function, then

THEOREM 7.2. *For each* $0 \leqq c \leqq \infty$, $\beta_R(AM; a) = \Phi(k_\alpha + 3c^2a(1 - a))$ *iff* $\lim_n (n^{\frac{1}{2}} d_n) = c$.

PROOF. It is easy to see that the test based on R is monotone (e.g. [29] and [28], p. 480). Thus Theorem 7.1 shows that one needs to compute

$\lim_n \inf [\beta_R(B(r, d_n, e; \cdot, \cdot)); 0 \leqq r \leqq 1 - d, e > 0]$. Theorem 1 of Bhuchongkul (1964) on uniform convergence to normality yields

$$|\inf \beta_R(B(r, d_n, e; \cdot, \cdot)) - \inf \Phi(k_\alpha + [E(r, d_n, e) - E_0]/\sigma_0)| \to 0 \quad \text{as} \quad n \to \infty,$$

where $E(r, d_n, e) = \int_0^1 \int_0^1 xF_Y(y)\, dB(r, d_n, e; x, y)$, $E_0 = \frac{1}{4}$ and $\sigma_0 = 1/12n^{\frac{1}{2}}$. Upon computing $F_Y(y) = ay + (1 - a)B_1(r, d_n, e; y)$ and $\lim_{e \to 0^+} E(r, d_n, e) = [1 + a(1 - a)\, d_n^2]/4$, the result follows.

Let $\sigma^2(J) = \int_0^1 J^2(u)\, du - (\int_0^1 J(u)\, du)^2$ and $A(a; J) = [a \int_0^1 J(u)\, du - \int_0^a J(u)\, du]/\sigma(J)$; then the following holds for the class $\mathfrak{J}$ of Bhuchongkul statistics defined in Section 6.

LEMMA 7.3. *For each* $T = T(J_n, L_n)$ *in* $\mathfrak{J}$ *and each* $0 \leqq c \leqq \infty$, *one has*

(i) *if the test based on* T *is monotone and* $L(x) = \lim_n L_n(x)$ *is of the form* $c_1 x + c_2$, *then* $\beta_T(AM; a) = \Phi(k_\alpha + 3^{\frac{1}{2}}c^2 A(a; J))$ *iff* $\lim_n (n^{\frac{1}{4}} d_n) = c$;

(ii) *if* $L(x)$ *is not of the form* $c_1 x + c_2$, *then* $\hat{\beta}_T(AM; a) < \Phi(k_\alpha + 3^{\frac{1}{2}}c^2 A(a; J))$ *whenever* $\lim_n (n^{\frac{1}{4}} d_n) = c$;

(iii) *in all cases,* $\hat{\beta}_T(AM; a) \leqq \Phi(k_\alpha + 3^{\frac{1}{2}}c^2 A(a; J))$ *whenever* $\lim_n (n^{\frac{1}{4}} d_n) = c$.

PROOF. (i) follows from the arguments of Theorem 7.2. Since $B(r, d, e; \cdot, \cdot)$ is in $\Omega(a; d)$, then (ii) will follow if one can show that there exists $0 < r < 1$ and $e > 0$ such that $\lim_n \beta_T(B(r, d_n, e; \cdot, \cdot)) < \Phi(k_\alpha + 3^{\frac{1}{2}}c^2 A(a, J))$. Let $B(r, d, e; \cdot, \cdot)$ be the alternative for which $B_1(r, d, e; t)$ equals $B_1(r, d; t)$ for t outside $(u + d, u + d + e)$ and equals the line $[(d/e) + 1]t - d[1 + (u + d)/e]$ inside this interval. Then lengthy computations yield

$$\lim_{e \to 0^+} \lim_n \{E[T(J, L) | B(r, d, e; \cdot, \cdot)] - E[T(T(J, L)) | H_0]\}/\sigma[T(J, L) | H_0] = A(a; J)L'(r)/2\sigma(L)$$

where $\sigma^2(L) = \int_0^1 L^2(u)\, du - (\int_0^1 L(u)\, du)^2$.

In the proof of Lemma 2.2 of [13] it is shown that there exists $0 < r < 1$ such that $L'(r)/\sigma(L) < 2 \cdot 3^{\frac{1}{2}}$.

(ii) now follows from Theorem 1 of Bhuchongkul (1964). (iii) follows by the same argument.

Let $A(J)$ denote $A(\frac{1}{2}; J)$, then

LEMMA 7.4. $\sup A(J) = \frac{1}{2}$, *where the* sup *is over the class of symmetrical* J *for which* $0 < \int_0^1 J^2(u)\, du < \infty$.

PROOF. Without loss of generality, assume that $\int_0^1 J(u)\, du = 0$; then one can write $4A^2(J) = (\int_{\frac{1}{2}}^1 J(u) 2\, du)^2 / \int_{\frac{1}{2}}^1 J^2(u) 2\, du$. From this and the elementary inequality $EX^2 \geqq (EX)^2$, one immediately obtains $4A^2(J) \leqq 1$ and thus $A(J) \leqq \frac{1}{2}$.

To see that the upper bound is attained, define $J_0(u) = 1$ for u in $(\frac{1}{2}, 1]$; $J_0(\frac{1}{2}) = 0$; and $J_0(u) = -1$ for u in $[0, \frac{1}{2})$. J_0 does not satisfy the conditions of [7], but there exists a sequence $\{J^k\}$ satisfying these conditions whose limit is J_0 and for which $\lim_k A(J^k) = A(J_0)$.

From the above proof, it is easily seen that there does not exist J such that (i) $A(J) = \frac{1}{2}$; and (ii) $T(J, L)$ is in $\mathfrak{J}(S)$ (of Section 6); hence Lemma 7.3 (iii) and Lemma 7.4 yield

COROLLARY 7.1. *If T is in $\mathfrak{I}(S)$, then* $\hat{\beta}_T(AM) < \Phi(k_\alpha + 3^{\frac{1}{2}}c^2/2)$.

Let $W = n^{-1}(n+1)^{-1} \sum J_0(r_i/(n+1))s_i$, then

LEMMA 7.5. *If* $E_H(W) = \iint J_0(x)y\, dH(x, y)$, *$B$ denotes* $B(r, d_n, e; \cdot, \cdot)$ *and* $\lim_n d_n = 0$, *then* $\lim_n P((12n)^{\frac{1}{2}}[W - E_B(W)] \leqq t \mid B) = \Phi(t)$ *uniformly with respect to t, r and e.*

PROOF. Let $T(J^k) = n^{-1}(n+1)^{-1} \sum J^k(r_i/(n+1))s_i$ be such that (i) J^k equals J_0 outside the interval $(\frac{1}{2} - k^{-1}, \frac{1}{2} + k^{-1})$; (ii) $J^k(u) = -J^k(1-u)$ for each u in $(0, 1)$; and (iii) $T(J^k)$ is in $\mathfrak{I}(S)$. Thus $T(J^k)$ is a sequence of statistics in $\mathfrak{I}(S)$ converging to $W = T(J_0)$. From the uniformity arguments of Chernoff and Savage (1958) and Bhuchongkul (1964), it follows that

$$P(n^{\frac{1}{2}}D_k^{-1}[T(J^k) - E_B(T(J^k))] \leqq t \mid B)$$

tends to $\Phi(t)$ uniformly in t, r, e and $k \geqq 4$ provided d_n tends to zero; where $D_k = \sigma(J^k)/(12)^{\frac{1}{2}}$ and $E_B(T(J^k)) = \iint J^k(x)y\, dB(x, y)$. One finds that D_k tends to 1 and $E_B(T(J^k))$ tends to $E_B(T(J_0)) = E_B(W)$ as k tends to infinity. Thus, upon taking limits as k tends to infinity, the result follows.

COROLLARY 7.2. $\beta_W(AM) = \Phi(k_\alpha + 3^{\frac{1}{2}}c^2/2)$.

PROOF. When n is even and the data have been permuted to make the x's ordered, then W is equivalent to $\sum_{\frac{1}{2}n+1}^{n} s_i$ and thus $\sum \epsilon(y_i - y_j)$, where the summation is over $i = \frac{1}{2}n + 1, \cdots, n$ and $j = 1, \cdots, \frac{1}{2}n$. It follows that the test based on W is monotone. A similar argument holds when n is odd. Using Lemma 7.5 and the arguments of Theorem 7.2 and Lemma 7.3, the result follows.

REFERENCES

[1] BELL, C. B. (1964). A characterization of multisample distribution-free statistics. *Ann. Math. Statist.* **35** 735–738.

[2] BELL, C. B. and BELLOT, F. (1965). Nota sobre los estadisticos no parametricos de Pitman. *Trabajos estadist.* **16** 25–37.

[3] BELL, C. B., BLACKWELL, D. and BREIMAN, L. (1960). On the completeness of order statistics. *Ann. Math. Statist.* **31** 794–797.

[4] BELL, C. B. and DOKSUM, K. A. (1965). Some new distribution-free statistics. *Ann. Math. Statist.* **36** 203–214.

[5] BELL, C. B. and DOKSUM, K. A. (1966). "Optimal" one-sample distribution-free tests and their two-sample extensions. *Ann. Math. Statist.* **37** 120–132.

[6] BELL, C. B., MOSER, J. M. and THOMPSON, R. (1966). Goodness criteria for two sample distribution-free tests. *Ann. Math. Statist.* **37** 133–142.

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

[8] BIRNBAUM, Z. W. (1953). On the power of a one-sided test of fit for continuous probability functions. *Ann. Math Statist.* **24** 484–489.

[9] BIRNBAUM, Z. W. and TANG, V. K. T. (1964). Two simple distribution free tests of goodness of fit. *Rev. Intern. Statist. Inst.* **32** 2–13.

[10] CAPON, J. (1961). Asymptotic efficiency of certain locally most powerful rank tests. *Ann. Math. Statist.* **32** 88–100.

[11] CHAPMAN, D. G. (1958). A comparative study of several one-sided goodness-of-fit tests. *Ann. Math. Statist.* **29** 655–674.

[12] CHERNOFF, H. and SAVAGE, L. R. (1958). Asymptotic normality and efficiency of certain nonparametric tests. *Ann. Math. Statist.* **29** 972–994.

[13] DOKSUM, K. A. (1966). Asymptotically minimax distribution free procedures. *Ann. Math. Statist.* **37** 619–628.

[14] DURBIN, J. (1961). Some methods of constructing exact tests. *Biometrika.* **48** 41–55.
[15] EHRENBERG, A. S. S. (1951). Note on normal transformation of ranks. *British J. Psych.* **4** 133–134.
[16] HÁJEK, J. (1962). Asymptotically most powerful rank-order tests. *Ann. Math. Statist.* **33** 1124–1147.
[17] JOGDEO, K. (1964). Nonparametric methods for regression. *Mathematisch Centrum, Amsterdam. Report* S 330.
[18] KONIJN, H. S. (1956). On the power of certain tests for independence in bivariate populations. *Ann. Math. Statist.* **26** 300–323.
[19] LE CAM, L. (1960). Locally asymptotically normal families of distributions. Univ. California Publ. *Statist.* **3** 37–98.
[20] LEHMANN, E. L. (1959). *Testing Statistical Hypotheses.* Wiley, New York.
[21] LEHMANN, E. L. and STEIN, C. (1949). On the theory of some nonparametric hypotheses. *Ann. Math. Statist.* **20** 28–45.
[22] PITMAN, E. J. G. (1937a). Significance tests which may be applied to samples from any population. *Suppl. J. Roy. Statist. Soc.* **4** 119–130.
[23] PITMAN, E. J. G. (1937b). Significance tests which may be applied to samples from any population. II *Suppl. J. Roy. Statist. Soc.* **4** 225–232.
[24] PITMAN, E. J. G. (1938). Significance tests which may be applied to samples from any population. III. *Biometrika.* **29** 322–335.
[25] RAO, C. R. and POTI, S. J. (1946). On locally most powerful tests when alternatives are one sided. *Sankhyā.* **7** 439–440.
[26] SCHEFFÉ, H. (1943). On a measure problem arising in the theory of non-parametric tests. *Ann. Math. Statist.* **14** 227–233.
[27] HOEFFDING, W. (1953). On the distribution of the expected values of the order statistics. *Ann. Math. Statist.* **24** 93–100.
[28] KENDALL, M. G. and STUART, A. (1961). *The Advanced Theory of Statistics*, **2**. Griffin, London.
[29] LEHMANN, E. L. (1966). Some concepts of dependence. *Ann. Math. Statist.* **37** 1137–1153.