"OPTIMAL" ONE-SAMPLE DISTRIBUTION-FREE TESTS AND THEIR TWO-SAMPLE EXTENSIONS[1]

By C. B. Bell and K. A. Doksum

San Diego State College and University of California, Berkeley

1. Introduction and summary. The object of this paper is the development of a theory of optimal one-sample goodness-of-fit tests and of optimal two-sample randomized distribution-free (DF) statistics analogous to the well-known results of Hoeffding (1951), Terry (1952), Lehmann (1953), (1959), Chernoff and Savage (1958), Capon (1961) and others for two-sample nonrandomized rank statistics. For $Y_1, \cdots, Y_n$ a random sample from a population with continuous distribution function G, one tests in the one-sample case $H_0 : G = F$ vs. $H_1 : G \neq F$, where F is some known continuous distribution function. From the Neyman-Pearson lemma, distribution-free tests that are most powerful (MP) for any H vs. K satisfying $KH^{-1} = GF^{-1}$, are obtained. From these MP distribution-free tests, one can on paralleling the derivations ([14], [25], [17], [18], [7]) for locally MP tests in the two-sample case obtain locally MP tests in the one-sample case. Further, it is found that the class of alternatives, for which a critical region of the form $[\sum J[F(y_i)] > c]$ is locally MP, is the class of G's that consists of "contaminated" Koopman-Pitman distributions as given in Section 5.

Randomized versions of the two-sample MP and locally MP rank statistics are considered and shown to be asymptotically equivalent to the locally MP rank statistics.

2. Preliminaries. To insure the existence of the appropriate derivatives and inverses it is necessary to restrict attention to distribution functions from the Scheffé (1943) class Ω_3^* of strictly increasing absolutely continuous distributions on R_1 and to its symmetrically complete sublcasses [24], [1].

In view of the structure and characterization results of Birnbaum and Rubin (1954) and Bell (1964a), it will be feasible to consider in the one-sample case only statistics of the form $\psi[F(Y_1), \cdots, F(Y_n)]$, i.e. statistics of *structure* (d); and in the two-sample case only *rank statistics*; i.e. statistics depending only on the ranks $R(X_1), \cdots, R(X_m)$; $R(Y_1), \cdots, R(Y_n)$ in the combined sample $X_1, \cdots, X_m$; $Y_1, \cdots, Y_n$. One should note that this class of rank statistics does not include the usual Pitman (1939) conditional statistics; and that the characterization of the strongly distribution-free statistics as rank statistics requires a null boundary condition of Scheffé (1943).

Note that the difference between the one and two-sample problem is that in the one-sample case F is known whereas in the two-sample case F is unknown while a random sample $X_1, \cdots, X_m$ from F is given. See Moses (1964).

Received 18 June 1964; revised 9 June 1965.

[1] This research was supported by the National Science Foundation through grant NSFG-25220 and the General Electric Corporation.

It is known [6], [2], [3] that the structure (d) and rank statistics are *strongly distribution-free* in the sense that if $\beta(F, G)$ denotes the probability that a test of H_0 vs. H_1 based on one of these statistics rejects H_0, then $\beta(F, G)$ is a function of (F, G) only through $GF^{-1} = (FG^{-1})^{-1}$; and, hence, structure (d) and rank statistics are distribution-free. From the strongly distribution-free property, it follows that if $GF^{-1} = KJ^{-1}$, then $\beta(F, G) = \beta(J, K) = \beta(U, H)$, where U is the standard uniform distribution on the unit interval I and $H = GF^{-1}$, which is also a distribution function on I.

In some special cases, the derived statistics will be monotone and it will be feasible to make use of the results of Chapman (1958), Lehmann (1959), and Bell-Moser-Thompson (1965). In order to do this one needs the following definitions and notation:

(1) $G(\cdot\,; d)$ is that cpf on I such that $G(u; d) = 0$ for $0 < u < d$; $= u - d$ for $d \leqq u < 1$, $= 1$ for $u = 1$; where $0 < d < 1$.

(2) $G(\cdot\,; u_0, d)$ is that cpf on I such that $G(u; u_0, d) = u$ for $0 < u < u_0$ and $u_0 + d \leqq u \leqq 1$, $= u_0$ for $u_0 \leqq u < u_0 + d$; where $0 \leqq u_0 < u_0 + d \leqq 1$.

(3) A monotone statistic is such that

(i) in the one-sample case $T(y_1, \cdots, y_n) \leqq T(y_1^*, \cdots, y_n^*)$ whenever $y_i \leqq y_i^*$ for all i; and

(ii) in the two-sample case $S(x_1, \cdots, x_m; y_1, \cdots, y_n) \leqq S(x_1^*, \cdots, x_m^*; y_1^*, \cdots, y_n^*)$ whenever $x_i \leqq x_i^*$ and $y_i \geqq y_i^*$ for all i and j.

(4) $G(F, d) = \{G; G$ is a continuous distribution on R_1, $G < F$, and $\sup_x [F(x) - G(x)] = d\}$.

In terms of these concepts, it is known ([8], [18], p. 187, [5]) that if T is an arbitrary monotone structure (d) statistic or monotone rank statistic, then the critical region $[T > c]$ is unbiased for one-sided alternatives $G < F$, and the maximum and minimum power for alternatives H in the class $G(F, d)$ are given by

(i) $\sup \beta_T(F, H) = \beta(U, G(\cdot\,; d))$; and

(ii) $\inf \beta_T(F, H) = \inf \beta(U, G(\cdot\,; u_0, d))$, where the latter infimum is over the set $0 \leqq u_0 \leqq 1 - d$.

3. MP one-sample goodness-of-fit tests. Let $H = GF^{-1}$, then H is the distribution of $F(Y_i)$ $(i = 1, \cdots, n)$, thus the Neyman-Pearson lemma implies that if H has a density h, then the MP critical regions for $H_0 : G = F$ vs. $H_1 : G \neq F$ is of the form $\{\prod h[F(y_i)] > c\}$. In view of this and the results of Section 2, one has

THEOREM 3.1. *If H is a strictly increasing continuous distribution on I with a density h, if $T = T_H(F, y_1, \cdots, y_n) = \sum \ln h[F(y_i)]$, and if one considers critical regions $\{T > c\}$, then*

(i) *T has structure (d); is* SDF *and* DF;

(ii) *$\beta_T(F, G) = \beta_T(K, J)$ whenever $GF^{-1} = JK^{-1}$;*

(iii) *$\{T > c\}$ is* MP *for F* vs. *G whenever $GF^{-1} = H$;*

(iv) *$\{T > c\}$ is admissible for alternatives in Ω_3^*.*

Note that (iv) follows from (iii) and the uniqueness part of the Neyman-Pearson lemma ([18], p. 65).

In several cases of importance T is found to be a monotone function of the form $a \sum J[F(y_i)] + b$, where J is the inverse of some distribution function.

EXAMPLE 3.1. $-\Pi = 2 \sum \ln F(y_i)$; $\Pi' = -2 \sum \ln [1 - F(y_i)]$; $\bar{U} = n^{-1} \sum F(y_i)$; and $\bar{Z} = n^{-1} \sum \Phi^{-1}[F(y_i)]$, where Φ is the standard normal distribution. The J's here are, respectively, inverses of the negative exponential ($\exp(x)$, $x < 0$), exponential ($1 - \exp(-x)$, $x > 0$), standard uniform and normal distributions. The alternatives against which these tests are MP are given in Section 5.

Whenever T is a monotone one-sample statistic one can prove, using the results at the end of Section 2,

THEOREM 3.2. *If the h of Theorem 3.1 is a monotone increasing function, then for every F in Ω_3^* maximum and minimum power are given by*

(i) $\sup \beta_T(F, H) = \beta(U, G(\cdot, d))$;

(ii) $\inf \beta_T(F, H) = \inf \beta(U, G(\cdot; u_0, d))$; *where the supremum and first infimum are taken over the class $G(F, d)$ of one-sided alternatives, and the latter infimum is taken over the set $0 \leqq u_0 \leqq 1 - d$.*

Finally, from elementary considerations one gets

THEOREM 3.3. *If $E_F |\ln h[F(Y)]|$ and $E_G |\ln h[F(Y)]|$ are finite and if $E_F\{\ln h[F(Y)]\} < E_G\{\ln h[(F(Y)]\}$ at the alternative $G \neq F$, then the test based on T is consistent for the alternative G.*

In several important cases the MP test statistic is unwieldy and one would wish to look for approximations as in the following example:

EXAMPLE 3.2. Consider the problem of obtaining a distribution-free statistic which is MP for F and G whenever $GF^{-1} = F_\theta F_0^{-1}$, where $F_0(x) = 1/[1 + \exp(-x)]$ and $F_\theta(y) = 1/[1 + \exp(-y + \theta)]$, i.e. logistic translation alternatives. On employing Theorem 3.1, one finds that the MP statistic is of the form $T(n, \theta) = \sum \ln h_\theta[F(y_i)] = n\theta - 2 \sum \ln \{F_0(y_i) + [1 - F_0(y_i)] \exp(\theta)\}$. Since $T(n, \theta)$ is somewhat involved, it seems worthwhile to consider its series expansion about $\theta = 0$:

$$T(n, \theta) = 2\theta n(\bar{U} - \tfrac{1}{2}) + \theta^2 n[n^{-1} \sum F_0^2(y_i) - \bar{U}] + o(\theta^2) \tag{3.1}$$

where $\bar{U} = n^{-1} \sum F_0(y_i)$. This expansion suggests the use of the $\bar{U}$ statistic for values of θ near 0 since for these θ, $\bar{U}$ is approximately equivalent to the MP statistic $T(n, \theta)$.

More generally, the above example suggests that when there is sufficient regularity to guarantee the necessary partial expansion, it will be possible to develop derivations analogous to those for optimal two-sample statistics, e.g. Capon (1961). This is done in the next section.

4. Almost locally MP goodness-of-fit tests.

Since the development will be parallel to that of Capon (1961), the following regularity conditions of Capon will be required. For testing $H_0 : G = F$ vs. $H_1 : G = G_\theta$, and for $H_\theta = G_\theta F^{-1}$, we assume:

(4.1) The distribution H_θ on the unit interval I has a derivative h_θ, which along with $(\partial/\partial\theta)h_\theta$ is continuous wrt θ in some non-degenerate closed interval $I(\theta_0)$ containing θ_0 for almost all u in I.

(4.2) There exist functions $M_0(u)$ and $M_1(u)$ integrable over (0, 1) and such that for all u in (0, 1), $0 < h_\theta(u) \leqq M_0(u)$, and $|H_\theta(u)| \leqq M_1(u)$ for θ in $I(\theta_0)$.

(4.3) $$E_{H_\theta} |(\partial/\partial\theta) \ln h_\theta(u) |_{\theta=\theta_0}|^2 < \infty$$

for θ in some nondegenerate closed interval containing θ_0.

For classes of alternatives satisfying (4.1) and (4.2) one now defines $T' = \sum (\partial/\partial\theta) \ln h_\theta[F(y_i)] |_{\theta=\theta_0}$, which (from expansions of the type (3.1)) should have optimal local properties. More precisely, for a given parametric family $\{H_\theta(F)\} = \{G_\theta F^{-1}(F)\} = \{G_\theta\}$ with $H_{\theta_0}(u) = u(0 < u < 1)$ and MP test statistic $T(\theta) = (\theta - \theta_0)^{-1} \sum \ln h_\theta\{F(y_i)\}$ one makes the following definitions. (Note the change in notation from earlier T.)

DEFINITION 4.1. A statistic T' is *almost locally* MP(LMP) if and only if for each positive ϵ and η, there exists a positive δ satisfying $P_\theta(|T' - T(\theta)| > \eta) < \epsilon$ whenever $|\theta - \theta_0| < \delta$; and

DEFINITION 4.2. A statistic T' is *asymptotically efficient* if and only if the Pitman asymptotic relative efficiency $A(T', T(\theta)) = 1$.

Here, Pitnam asymptotic relative efficiency (ARE) is taken to be as defined by (for instance) Hodges and Lehmann (1961). Chernoff and Savage (1958) and Capon (1961) have pointed out that if S is any statistic, then $A(S, T(\theta)) \leqq 1$. Thus if T' is asymptotically efficient, then $A(S, T') \leqq 1$ for any statistic S. In the notation of van Eeden (1963), an asymptotically efficient statistic would be called a "best" statistic. Note that asymptotic efficiency is a local property although its designation (which is consistent with common use) does not indicate this.

With these definitions and regularity conditions one can parallel the proofs of Capon to prove

THEOREM 4.1. *For* F vs. $G_\theta = H_\theta(F)$, *as well as for each* J vs. K_θ *satisfying* $K_\theta J^{-1} = H_\theta$, *and families* $\{H_\theta\}$ *satisfying* (4.1)–(4.3) *above, the strongly distribution-free statistic* $T' = \sum (\partial/\partial\theta) \ln h_\theta[F(y_i)] |_{\theta=\theta_0}$ *is* (a) *almost* LMP *and* (b) *asymptotically efficient.*

PROOF. (a) is immediate from the series expansion of $T(\theta)$ [see (3.1)] and Definition 4.1. (b) follows from the definition of Pitman efficiency [12] if one sets $(\theta - \theta_0) = cn^{-\frac{1}{2}}$ for some non-zero constant c; and notes that in this case $n^{-\frac{1}{2}}[T(\theta) - T']$ tends to zero in probability.

In the notation of Rao and Poti (1946) and Lehmann (1959), p. 342, the statistic T' would be called LMP. It is called almost LMP here because it is in general not MP for any parameter value no matter how close it is to θ_0. [See (3.1).] This is in contrast to the properties of discrete rank tests [e.g. Capon (1961)].

At this point it is clear that whenever $T(\theta)$ is MP for F vs. $H_\theta(F)$, T' is locally optimal for the class (of appropriately regular alternatives) $\{H_\theta(F)\}$. Hence, it is natural to ask the following questions:

(I) When are T and T' equivalent statistics?

(II) For which classes of alternatives is T' almost LMP?

A question similar to (II) has been considered by J. Neyman in the parametric case. Following his notation, the classes in (II) can be called the domain of optimality of T'.

Answers to the above questions are given in the next section.

5. Optimal classes for $T(\theta)$ and T'. With regard to question (I) of Section 4 one should consider

EXAMPLE 5.1. For F vs. $G_\theta = F^\theta$, which occurs, for example, when F and G_θ are negative exponentials, one has $H_\theta(u) = G_\theta F^{-1}(u) = u^\theta$; $(\theta - 1)T(\theta) = n \ln \theta + (\theta - 1) \sum \ln F(y_i)$ and $T' = n + \sum \ln F(y_i)$. Consequently, $T(\theta)$ and T' are both equivalent to the statistic $-\text{II}$ of Example 3.1.

In generalizing this result one solves the equation $\ln h_\theta(u) = aS(u) + b(\theta)$, and obtains

THEOREM 5.1. *$T(\theta)$ and T' are equivalent statistics, [in the sense that $T(\theta) = \alpha(\theta)T' + \beta(\theta)$ for nonvanishing $\alpha(\theta)$] whenever*

(i) $h_\theta(u) = \exp\{a(\theta)S(u) + b(\theta)\}$; *or equivalently,*

(ii) $G_\theta'(x) = \exp\{a(\theta)S(F(x)) + b(\theta) + \ln F'(x)\}$, *where $a(\theta)$ and $b(\theta)$ are differentiable and $a'(\theta)$ is nonvanishing;*

(iii) *when either* (i) *or* (ii) *holds, then $T(\theta)$ and T' are both equivalent to the statistic $\sum_{i=1}^n S[F(y_i)]$.*

The class in Theorem 5.1 (ii), is of course, a Koopman-Pitman class; and such classes play an important role in answering question (II) of Section 4.

DEFINITION 5.1. A *contaminated* Koopman-Pitman distribution function is a distribution function which can be written in the form $[a(\theta)K_\theta(x) + R(x, \theta)]$, where $a(\theta) \to a(\theta_0) = 1$ and $R(x, \theta) \to R(x, \theta_0) = 0$ as $\theta \to \theta_0$, and $K_\theta(x)$ is a Koopman-Pitman distribution.

The following result is an immediate consequence of Section 4.

THEOREM 5.2. *$T' = \sum (\partial/\partial\theta) \ln h[F(y_i)]\,|_{\theta=\theta_0} = \sum J[F(y_i)]$ (with critical region $\{T' > c\}$) is*

(i) MP *for the class of F and G_θ which satisfy $(d/du)G_\theta F^{-1}(u) = h_\theta(u; F, G_\theta) =$* $\exp[a(\theta)J(u) + b(\theta)]$ *or* $G_\theta'(y) = \exp\{a(\theta)J[F(y)] + b(\theta) + \ln F'(y)\}$ *for some functions $a(\theta)$ and $b(\theta)$; and*

(ii) *almost* LMP *for the contaminated Koopman-Pitman class of all F and G_θ that satisfy*

$$h_\theta(u; F, G_\theta) = \exp[a(\theta)J(u) + b(\theta) + Q(u, \theta)]$$

or

$$G_\theta'(y) = \exp\{a(\theta)J[F(y)] + b(\theta) + \ln F'(y) + Q[F(y), \theta]\}$$

where $a(\theta)$, $b(\theta)$ and $Q(u, \theta)$ are such that $a(\theta)$ and $b(\theta)$ are differentiable, $a'(\theta)$ is nonvanishing, and $Q(u, \theta) = o(\theta - \theta_0)$ for all u in I.

With reference to applications of Theorem 5.2 it is instructive to reconsider Example 3.2 in terms of a contaminated Koopman-Pitman class.

EXAMPLE 5.2. For the logistic translation situation in Example 3.2, $h_\theta(u; F_0, G_\theta) = \exp(\theta)/[u + (1-u)\exp(\theta)]^2$ and $\bar{U} = n^{-1}\sum F(y_i)$ is almost LMP. On the other hand, from Theorem 5.2 (i) one sees that $\bar{U}$ is MP for $h_\theta^*(u) = \exp[a(\theta)u + b(\theta)]$, and, in particular, for $\tilde{h}_\theta(u) = [\exp(\theta u) - 1]/[\exp(\theta) - 1]$. Since this is the case, $h_\theta(u, F, G_\theta)$ should be expressible in the form $\exp[a(\theta)u + b(\theta) + Q(u, \theta)]$, where $Q(u, \theta) = o(\theta)$ for all u in I. This is so for

$$a(\theta) = 2[\exp(\theta) - 1], \qquad b(\theta) = -2 - \theta \quad \text{and} \quad Q(u, \theta) = \textstyle\sum_{r=2}^{\infty} r^{-1}u[1 - \exp(-\theta)]^r.$$

One can now proceed to consider some extensions of these results to the two-sample case.

6. Randomized two-sample statistics. For two independent random samples $X_1, \cdots, X_m$ and $Y_1, \cdots, Y_n$ from populations with continuous distributions F and G respectively, one tests in the two-sample case $H_0 : F = G$ vs. $G = H_\theta(F)$ where H_θ equals the uniform distribution U_0 on the unit interval if and only if $\theta = \theta_0$. Let $N = m + n$ and let $z_{Ni} = 1$ if the ith observation in the ordered combined sample of X's and Y's is an X, $z_{Ni} = 0$ otherwise. Further, it will be necessary to consider the order statistics $U(1), \cdots, U(N)$ of a U_0-random sample $U_1, \cdots, U_N$ which is independent of $X_1, \cdots, X_m; Y_1, \cdots, Y_n$.

Before constructing the proposed new statistics it is worthwhile to recall that in view of the comments of Section 2 attention may be restricted to rank statistics and that when H_θ satisfies conditions (4.1) and (4.2), then

(a) the MP rank test statistics due to Hoeffding (1951) are of the form

$$V(\theta) = E[\prod \{h_\theta[U(i)]\}^{z_{Ni}}] = E[\prod h_\theta\{U[R(x_i)]\}]$$

and

(b) the generally more tractable LMP rank statistics are of the form

$$V' = \sum E\{(\partial/\partial\theta) \ln h_\theta[U(i)] |_{\theta=\theta_0}\} z_{Ni} = \sum E\{(\partial/\partial\theta) \ln h_\theta[U(R(x_i))] |_{\theta=\theta_0}\}.$$

In practice the MP test statistics are often quite cumbersome; and the $E\{(\partial/\partial\theta) \ln h_\theta[U(i)] |_{\theta=\theta_0}\}$ of the LMP test statistic are quite often difficult to compute; or are not tabulated and available to statisticians. Further, in both cases only at most $\binom{N}{n}$ significance levels are obtainable without randomization. If one is willing to randomize to obtain exact significance levels, it seems reasonable to consider randomization at an early stage in order to circumvent the tabulation problem if possible.

The method which immediately suggests itself is that of removing the expectation signs to obtain $S^*(\theta) = \prod h_\theta\{U[R(X_i)]\}$ or $\ln S^*(\theta) = \sum \ln h_\theta\{U[R(X_i)]\}$ and $\sum (\partial/\partial\theta) \ln h_\theta\{U[R(x_i)]\} |_{\theta=\theta_0}$. Experience [4] with special versions of these statistic has shown that Pitman ARE can be improved if one includes the order statistics $U[R(Y_1)], \cdots, U[R(Y_m)]$ in the formulas for the statistic. Therefore, in the sequel, attention will be centered on the statistics

(A) $S(\theta) = m^{-1} \sum \ln h_\theta\{U[R(X_i)]\} - n^{-1} \sum \ln h_\theta\{U[R(Y_i)]\}$; and

(B) $S' = m^{-1} \sum (\partial/\partial\theta) \ln h_\theta\{U[R(X_i)]\} |_{\theta=\theta_0} - n^{-1} \sum (\partial/\partial\theta) \ln h_\theta \cdot \{U[R(Y_i)]\} |_{\theta=\theta_0}$.

The idea of randomizing by using order statistics has also been used by Durbin (1961).

Special versions of $V(\theta)$, V', $S(\theta)$ and S' have been treated by several authors. For example,

EXAMPLE 6.1. *For* $H_0 : G = F$ vs. $H_1 : G = 1 - (1 - F)^\theta$, which is the case for exponentials $F(x) = 1 - \exp(-x)$ and $G(x) = 1 - \exp(-\theta x)$, one has

$$V(\theta) = \theta^m E[\prod \{1 - U[R(x_i)]\}^{\theta-1}];$$

$$V' = m + \sum E[\ln \{1 - U[R(x_i)]\}];$$

$$S(\theta) = (\theta - 1)[m^{-1} \sum \ln \{1 - U[R(x_i)]\} - n^{-1} \sum \ln \{1 - U[R(y_i)]\}$$

and

$$S' = (\theta - 1)^{-1} S(\theta) |_{\theta=1}.$$

About this example one should note that V has been treated in [17]; V' in [23], [9] and [7]; that $S(\theta)$ and S' are equivalent, and are treated in [4]. In this last reference S' is written $m^{-1} \sum Z[R(X_i)] - n^{-1} \sum Z[R(Y_j)]$, where $Z(1), \cdots, Z(N)$ are the order statistics of a random sample with common distribution $1 - \exp(-x)$. The equivalence of the two expressions follows from the fact $-2[\ln(1 - U)]$ has distribution $1 - \exp(-x)$, whenever U has cpf U_0.

As in the example above, S and S' can often be written as monotone functions of the form $m^{-1} \sum H^{-1}\{U[R(X_i)]\} - n^{-1} \sum H^{-1}\{U[R(Y_i)]\}$ where H^{-1} is the inverse of some distribution H. From Lehmann (1959), p. 187, one obtains

THEOREM 6.1. *If the h_θ of formula* (A) *is increasing, then the critical region* $\{S(\theta) < c\}$ *is unbiased for one-sided alternatives* $G < F$. *The same results hold for* $(\partial/\partial\theta) \ln h_\theta |_{\theta=\theta_0}$ *of formula* (B) *and the critical region* $\{S' < c\}$.

Asymptotic optimal properties of $S(\theta)$ and S' will be established in the next section.

7. Optimal properties of the randomized two-sample statistics. Essentially, it will be established in this section that the randomized statistics $S(\theta)$ and S' are in some asymptotic sense as good as $V(\theta)$ and V'. In order to accomplish this the regularity conditions (4.1) and (4.2) are used as well as

$$0 < \lim_N (m/n) = r < \infty; \tag{7.1}$$

further there exists a constant K such that

$$|(\partial^i/\partial u^i) J(u)| \leqq K[u(1 - u)]^{-i-\frac{1}{2}-\delta} \quad \text{for} \quad i = 0, 1, 2 \tag{7.2}$$

and some $\delta > 0$, where $J(u) = (\partial/\partial\theta) \ln h_\theta(u) |_{\theta=\theta_0}$.

The following result of Hoeffding [15] is also used.

LEMMA 7.1. *If $f(x)$ is convex, $g(x)$ is convex and nondecreasing (for $x \geqq A$ if $f(y) \geqq y$); and if $\int x\, dF(x)$, $\int f(x)\, dF(x)$, and $\int g(f(x))\, dF(x)$ all exist, then* $\lim N^{-1} \sum_{j=1}^{N} g[Ef(Z(j, N))] = \int g(f(x))\, dF(x)$, *where $Z(j, N)$ is the* jth *order statistic of the random sample $Z_1, \cdots, Z_N$ with common distribution F.*

In order to apply this result, the following conditions are needed for the function $J(u) = (\partial/\partial\theta) \ln h_\theta(u) |_{\theta=\theta_0}$.

J satisfies $\int_0^1 J^2(u)\, du < \infty$ and at least one of the following conditions:

(7.3)
(a) J is the inverse of a distribution function H,
(b) J is convex and is bounded below,
(c) there exists a continuous distribution H such that $J[H(x)]$ is convex, bounded below, and $|\int x\, dH(x)| < \infty$.

In order to make S' easily comparable with the LMP rank statistic, the LMP rank statistics will be written in the form

$$\begin{aligned} W' &= m^{-1} \textstyle\sum_{i=1}^{m} E(J\{U[R(X_i)]\}) - n^{-1} \sum_{i=1}^{n} E(J\{U[R(Y_i)]\}) \\ &= (\lambda_1\lambda_2 N)^{-1} \textstyle\sum_{i=1}^{N} \{E(J[U(i)])\}(z_{Ni} - \lambda_1) \end{aligned}$$

where $\lambda_1 = m/N$ and $\lambda_2 = n/N$. W' is easily seen to be equivalent to V', thus W' is a LMP rank statistic whenever V' is.

The fact that S' is asymptotically equivalent to a LMP rank statistic is reflected in the following theorem:

THEOREM 7.1. *If H_θ is such that W' is defined, then*
(i) $E_\theta(S') = E_\theta(W')$. *If further* (7.1) *and* (7.3) *are satisfied, then*
(ii) $\mathrm{Var}_{\theta_0} [N^{\frac{1}{2}}(S' - W')] \to 0$ *as* $N \to \infty$, *and*
(iii) $\mathrm{Var}_\theta [N^{\frac{1}{2}}(S' - W')] \to 0$ *as* $N \to \infty$ *whenever either*
(a) $A_N = o(N)$ *as* $N \to \infty$ *or* (b) $A_N \leqq 0$, *where*

$$A_N = \textstyle\sum_{i<j} \mathrm{Cov}\, \{J[U(i)], J[U(j)]\} E_\theta(z_{Ni} - \lambda_1)(z_{Nj} - \lambda_1).$$

PROOF. Note that S' can be written $S' = (\lambda_1\lambda_2 N)^{-1} \sum_{i=1}^{N} J[U(i)](z_{Ni} - \lambda_1)$, then (i) is immediate since $U(i)$ and z_{Ni} are independent. Also note that

$$\mathrm{Var}_\theta [N^{\frac{1}{2}}(S' - W')] = [\lambda_1\lambda_2]^{-2} N^{-1} \textstyle\sum_{i=1}^{N} \{\mathrm{Var}\,(J[U(i)])\} E_\theta(z_{Ni} - \lambda_1)^2 + 2[\lambda_1\lambda_2]^{-2} N^{-1} A_N .$$

On applying Lemma 7.1, (iii) now follows since $E_\theta(z_{Ni} - \lambda_1)^2$ is bounded. (ii) follows from (iii) on observing that $E_{\theta_0}(z_{Ni} - \lambda_1)(z_{Nj} - \lambda_1) = -\lambda_1\lambda_2(N - 1)^{-1}$ for $i \neq j$.

The notion of asymptotic efficiency will be in the two-sample as in the one-sample case. Let $W' = (\lambda_1\lambda_2 N)^{-1} \sum \{E(J[U(i)])\}(z_{Ni} - \lambda_1)$ be a LMP rank statistic for $F = G$ vs. $G = H_\theta(F)$, $\theta \neq \theta_0$, then

DEFINITION 7.1. Let $\theta = \theta_N$ be such that $\theta_N \to \theta_0$ as $N \to \infty$ and such that $N^{\frac{1}{2}}[W' - E(W')]$ converges in law (under P_{θ_N}) to a nondegenerate random variable, then S' is an *asymptotically* LMP *rank* (ALMPR) *statistic* for $F = G$ vs. $G = H_\theta(F)$ $(\theta \neq \theta_0)$, if and only if for $\epsilon > 0$, $P_{\theta_N}(N^{\frac{1}{2}} |S' - W'| \geqq \epsilon) \to 0$ as $N \to \infty$.

Note that the sequences $\{\theta_N\}$ of Definition 7.1 are often of the form $\{\theta_0 + c/N^{\frac{1}{2}}\}$. They will be of this form under the conditions given in the next result.

In order to apply the results of Hájek and LeCam [11] and Matthes and Truax (1965) on contiguity, the following condition will be needed. U is a random variable with distribution U_0.

$$\lim_{\theta\to\theta_0} E\{[(h_\theta^{\frac{1}{2}}(U) - h_{\theta_0}^{\frac{1}{2}}(U))/\theta h_{\theta_0}^{\frac{1}{2}}(U)] - J(U)\} = 0. \tag{7.4}$$

Using theorem 7.1, one can now conclude

COROLLARY 7.1. *If the conditions* (7.1), (7.3), *and* (7.4) *hold, then for testing* $F = G$ vs. $G = H_\theta(F)$, $\theta \neq \theta_0$,

(i) S *and* S' *are* ALMPR *statistics. If further* (4.1), (4.2) *and* (7.2) *hold, then*

(ii) S *and* S' *are asymptotically efficient.*

PROOF. For (i), note that Capon (1961) proved that W' is a LMP rank statistic. Thus Theorem 7.1 (ii) and the results of Hájek and LeCam [11] and Matthes and Truax (1965) imply that S' is an ALMPR statistic. (See Theorem 3 of Matthes and Truax.) Similarly, $S(\theta)$ is an ALMPR statistic since $S(\theta_0)$ is equivalent to S'.

For (ii), note that Capon (1961) proved that W' is asymptotically efficient and apply the same arguments as above.

Finally, one can establish the results analogous to those of Section 5. They are:

THEOREM 7.2. (a) $S(\theta)$ *and* S' *are equivalent statistics whenever* $h_\theta(u) = \exp\{a(\theta)J(u) + b(\theta)\}$ *or* $G_\theta'(x) = \exp\{a(\theta)J[F(x)] + b(\theta) + \ln F'(x)\}$*; and*

(b) *whenever* h_θ *is of the form in* (a) *both* $S(\theta)$ *and* S' *are* ALMPR *statistics, asymptotically efficient, and are both equivalent to* $m^{-1}\sum J\{U[R(X_i)]\} - n^{-1}\sum J\{U[R(Y_j)]\}$.

THEOREM 7.3. *Statistics of the form* $m^{-1}\sum J\{U[R(X_i)]\} - n^{-1}\sum J\{U[R(Y_j)]\}$ *are* ALMPR *statistics and asymptotically efficient for the contaminated Koopman-Pitman class*

$$h_\theta(u) = \exp\{a(\theta)J(u) + b(\theta) + Q(u, \theta)\} \text{ or}$$

$$G_\theta'(x) = \exp\{a(\theta)J[F(x)] + b(\theta) + \ln F'(x) + Q(F(x), \theta)\},$$

where $Q(u, \theta) = o(\theta - \theta_0)$ *for all* u.

The parallelism between the various one-and two-sample statistics is further indicated in the following section.

8. Applications. The results of the preceding sections will now be applied to special alternatives and statistics. In both the one-sample and two-sample case the alternatives will be denoted by $G_\theta = H_\theta(F)$ where there exist some θ_0 such that $h_\theta(u) = u$ for each u in [0, 1] if and only if $\theta = \theta_0$. Let J_0 be a function on the unit interval (possibly unrelated to h_θ). Consider the one-sample statistic $T(J_0) = \sum J_0[F(Y_i)]$ and the two-sample statistics

$$W(J_0) = m^{-1}\sum E[J_0(U[R(X_i)])] - n^{-1}E[J_0(U[R(Y_i)])]$$

and $S(J_0) = m^{-1}\sum J_0(U[R(X_i)]) - n^{-1}\sum J_0(U[R(Y_i)])$. $T(J_0)$ is MP for F vs. $G_\theta = H_\theta(F)$ if $J_0 = \ln h_\theta$ and is almost LMP and asymptotically efficient for F vs. $G_\theta = h_\theta(F)$ if $J_0 = (\partial/\partial\theta) \ln h_\theta |_{\theta=\theta_0}$. Similar statements hold for $W(J_0)$ and $S(J_0)$ in the two-sample case.

The asymptotic relative efficiency (ARE) of $W(J_0)$ wrt Student's two-sample statistic t_2 has been studied by Hodges and Lehmann (1956), Chernoff and Savage (1958) and others for special cases of J_0 and various alternatives. Let t_1 denote the one-sample Student statistic, then the result $A(T(J_0), t_1) = A(W(J_0), t_2) = A(S(J_0), t_2)$ (for alternatives $H_\theta(F)$ such that these ARE's are computable) follows from arguments of the preceding sections.

The above notation and result will now be used in the following examples and table; moreover Φ will denote the standard normal distribution. The results in the examples hold for a general normal distribution after obvious modifications. Finally, let $Q(u, \theta)$ be such that $Q(u, \theta) = o(\theta - \theta_0)$ as $\theta \to \theta_0$.

EXAMPLE 8.1. (Normal translations). For Φ vs. the alternatives $G_\theta(x) = H_\theta[\Phi(x)] = \Phi(x - \theta)$ one finds $H_\theta(u) = \Phi(\Phi^{-1}(u) - \theta))$, $\ln h_\theta(u) = \theta\Phi^{-1}(u) + \frac{1}{2}\theta^2$ and $(\partial/\partial\theta) \ln h_\theta(u) |_{\theta=0} = \Phi^{-1}(u)$. Thus $\bar{Z}_1 = n^{-1}\sum \Phi^{-1}[F(Y_i)]$ is uniformly MP in the one-sample case (Theorems 5.1 and 5.2) and $\bar{Z}_2 = n^{-1}\sum \Phi^{-1}(U[R(X_i)]) - m^{-1}\sum \Phi^{-1}(U[R(Y_i)])$ is an ALMPR statistic and is asymptotically efficient in the two-sample case for Φ vs. $\Phi(x - \theta)$ (Corollary 7.1).

Similarly, $\bar{Z}_1$ is uniformly most powerful for any continuous F vs. $G_\theta(x) = \Phi\{\Phi^{-1}[F(x) - a(\theta)]\}$, $\theta \neq \theta_0$, where $a(\theta) = 0$ if and only if $\theta = \theta_0$. Further, $\bar{Z}_1$ and $\bar{Z}_2$ are asymptotically efficient for F vs. $\Phi\{\Phi^{-1}(F(x)) - a(\theta) + Q(F(x), \theta)\}$, where $a'(\theta) \neq 0$ (Theorems 5.2 and 7.3).

From the results of Chernoff and Savage (1958) and the remark preceding this example, it follows that $A(\bar{Z}_1, t_1) = A(\bar{Z}_2, t_2) \geqq 1$ for F vs. $F(x - \theta)$ (e.g. translation) with equality if and only if F is a normal cpf.

Finally, $\bar{Z}_1$ and $\bar{Z}_2$ have the desirable property that their null-distributions ($F = G$) are normal with means zero and variances n^{-1} and $n^{-1} + m^{-1}$ respectively (see [4]).

EXAMPLE 8.2. (Normal scale). Φ vs. $\Phi(x/\theta) = H_\theta[\Phi(x)]$ yields $H_\theta(u) = \Phi(\Phi^{-1}(u)/\theta)$, $\ln h_\theta(u) = \frac{1}{2}[\Phi^{-1}(u)]^2(1 - \theta^{-2}) - \ln \theta$, $(\partial/\partial\theta) \ln h_\theta(u) |_{\theta=1} = [\Phi^{-1}(u)]^2 - 1$. Hence $Z_1^2 = \sum \{\Phi^{-1}[F(Y_i)]\}^2$ is uniformly MP in the one-sample case (Theorems 5.1 and 5.2) and

$$Z_2^2 = m^{-1}\sum \{\Phi^{-1}(U[R(X_i)])\}^2 - n^{-1}\sum \{\Phi'(U[R(Y_i)])\}^2$$

is an ALMPR statistic and is asymptotically efficient in the two-sample case (Corollary 7.1).

Moreover, Z_1^2 is uniformly MP for F vs. $G_\theta(x) = \Phi\{\Phi^{-1}[F(x)]/b(\theta)\}$, $\theta \neq \theta_0$, where $b(\theta) = 1$ if and only if $\theta = \theta_0$; and Z_1^2 and Z_2^2 are asymptotically efficient for F vs. $\Phi[\Phi^{-1}[F(x)]/b(\theta) + Q(F(x), \theta))$ (Theorems 5.2 and 7.3).

The null-distribution of Z_1^2 is χ_n^2; if $m = n$, then nZ_2^2 is distributed as the difference of two independent chi-square variates, each with n degrees of freedom. This latter distribution is discussed in [4], Table 5.1.

EXAMPLE 8.3. (Normal translation and scale). As pointed out by Kendall and Stuart (1961) and others, it is desirable to have statistics powerful for alternatives that are a combination of location and scale. Consider therefore the following:

Let $a(\theta)$ be such that $a(\theta) = 1$ if and only if $\theta = 0$ and such that $a'(\theta) \neq 0$. For testing Φ vs. $G_\theta(x) = \Phi(a(\theta)[x - \theta])$, $\theta \neq 0$, one obtains $\ln h_\theta(u) = \ln a(\theta) + \frac{1}{2}\{[1 - a^2(\theta)]\{\Phi^{-1}(u)\}^2 + 2\theta a^2(\theta)\Phi^{-1}(u) - \theta^2 a^2(\theta)\}$ and

$$(\partial/\partial\theta) \ln h_\theta(u) |_{\theta=0} = \Phi^{-1}(u) - a'(0)\{\Phi^{-1}(u)\}^2.$$

Thus the MP statistic depends on θ while the asymptotically efficient statistics are $Z_1 = \bar{Z}_1 - n^{-1}a'(0)Z_1^2$ and $Z_2 = \bar{Z}_2 - a'(0)Z_2^2$ where $\bar{Z}_1$, $\bar{Z}_2$, Z_1^2 and Z_2^2 are as in Examples 8.1 and 8.2. Z_1 and Z_2 are also asymptotically efficient for any continuous F vs. $G_\theta(x) = \Phi(a(\theta)[\Phi^{-1}(F(x)) - \theta] + Q(F(x), \theta))$.

The null-distributions of Z_1 and Z_2 are asymptotically normal. Their means are $-a'(0)n - 1$ and 0 and the variances are $2n^{-1}\{a'(0)\}^2$ and $2(n^{-1} + m^{-1})\{a'(0)\}^2$, respectively. Note that when $a(\theta) = 1/(\theta + 1)$, then $a'(0) = -1$, thus $Z_1 = \bar{Z}_1 + n^{-1}Z_1^2$ and $Z_2 = \bar{Z}_2 + Z_2^2$ for this case.

EXAMPLE 8.4. (Normal contamination). Let $0 \leqq a(\theta) \leqq 1$ be a differentiable function possible constant and such that $0 < a(1) = c \leqq 1$, then for the contamination alternatives Φ vs. $G_\theta(x) = a(\theta)\Phi(x) + (1 - a(\theta))\Phi(x/\theta) = H_\theta(\Phi(x))$ one has that the MP statistic depends on θ while $(\partial/\partial\theta) \ln h_\theta(u) |_{\theta=1} = (c - 1) + (1 - c)\{\Phi^{-1}(u)\}^2$. Thus the asymptotically efficient statistics are Z_1^2 and Z_2^2 of Example 8 .2.

The derivations in the above examples can also be carried out for cpf's other than Φ. Let F be a continuous cpf, then the following table gives in the first column hypothesis, in the second column alternatives; from the third column can be obtained the MP one-sample statistic $T(J_0)$, and from the fourth column can be read the asymptotically efficient statistics $T(J_1)$ and $S(J_1)$. a and b are convenient constants that may differ from column to column and from row to row.

TABLE 8.1

Hyp	Alternatives	$J_0(u) = a \ln h_\theta(u) + b$	$J_1(u) = a(\delta/\delta\theta) \ln h_\theta(u) \mid_{\theta=\theta_0} + b$
$F = K$	$\exp(x\theta)$	$\ln u$	$\ln u$
$F = C$	$C(x - \theta)$	$\ln [C'(C^{-1}(u) - \theta)/C'(C^{-1}(u))]$	$C^{-1}(u)/(1 + [C^{-1}(u)]^2)$
$F = C$	$C(x/\theta)$	$\ln [C'(\theta^{-1}C^{-1}(u))/C'(C^{-1}(u))]$	$[C^{-1}(u)]^2/(1 + [C^{-1}(u)]^2)$
F	F^θ	$\ln u$	$\ln u$
F	$(1 - \theta)F + \theta F^2$	$\ln (1 - \theta + 2\theta u)$	u
F	$(1 - \theta)F + \theta F^k$	$\ln (1 - \theta + k\theta u^{(k-1)})$	$u^{(k-1)}$
F	$(e^{\theta F} - 1)(e^\theta - 1)^{-1}$	u	u
F	$F(x - \theta)$	$\ln [F'(F^{-1}(u) - \theta)/F'(F^{-1}(u))]$	$F''(F^{-1}(u))/F'(F^{-1}(u))$
F	$F(x/\theta)$	$\ln [F'(\theta^{-1}F^{-1}(u))/F'(F^{-1}(u))]$	$F''(F^{-1}(u))F^{-1}(u)/F'(F^{-1}(u))$
F	$F[F + e^\theta(1 - F)]^{-1}$	$\ln [u + e^\theta(1 - u)]$	u

It is not known whether $S(J_1)$ is asymptotically efficient in rows 2 and 3. Rows 8 and 9 only apply when F satisfies appropriate regularity conditions. C denotes the standard Cauchy distribution and K the exponential distribution defined by $K(x) = \exp(x)$ if $x \leqq 0$ and $K(x) = 1$ if $x > 0$.

REFERENCES

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

[2] BELL, C. B. (1964a). Some basic theorems of distribution-free statistics. *Ann. Math. Statist.* **35** 150–156.

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

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

[5] BELL, C. B., MOSER, J. M. and THOMPSON, R. (1965). Goodness criteria for two sample distribution-free tests. *Ann. Math. Statist.* **37** 133–142.

[6] BIRNBAUM, Z. W. and RUBIN, H. (1954). On distribution-free statistics. *Ann. Math. Statist.* **25** 593–598.

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

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

[9] CHERNOFF, HERMAN and SAVAGE, I. RICHARD (1958). Asymptotic normality and efficiency of certain nonparametric tests. *Ann. Math. Statist.* **29** 972–994.

[10] DURBIN, J. (1961). Some methods of constructing exact tests. *Biometrika* **48** 41–55.

[11] HÁJEK, J. (1962). Asymptotically most powerful rank-order tests. *Ann. Math. Statist.* **33** 1124–1147.

[12] HODGES, J. L., JR. AND LEHMANN, E. L. (1956). The efficiency of some nonparametric competitors of the t-test. *Ann. Math. Statist.* **27** 324–335.

[13] HODGES, J. L., JR. and LEHMANN, E. L. (1961). Comparison of the normal scores and Wilcoxon tests. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **1** 307–318. Univ of California Press.

[14] HOEFFDING, W. (1951). "Optimum" nonparametric tests. *Proc. Second Berkeley Symp. Math. Statist. Prob.* Univ of California Press.

[15] HOEFFDING, W. (1953). On the distribution of the expected values of the order statistics. *Ann. Math. Statist.* **24** 93–100.

[16] KENDALL, M. G. and STUART, A. (1961). *The Advanced Theory of Statistics*, **2**. Griffin, London.

[17] LEHMANN, E. L. (1953). The power of rank tests. *Ann. Math. Statist.* **24** 23–43.

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

[19] MATTHES, T. K. and TRUAX, D. R. (1965). Optimal invariant rank tests for the k-sample problem. *Ann. Math. Statist.* **36** 1207–1222.

[20] MOSES, L. E. (1964). One sample limits of some two-sample rank tests. *J. Amer. Statist. Assoc.* **59** 645–651.

[21] PITMAN, E. J. G. (1939). Test of hypotheses concerning location and scale parameters. *Biometrika* **31** 200–215.

[22] RAO, C. R. and POTI, S. J. (1946). On locally most powerful tests when alternatives are one sided. *Sankhyā* **7** 439–440.

[23] SAVAGE, I. R. (1956). Contributions to the theory of rank order statistics. Two sample case. *Ann. Math. Statist.* **27** 590–616.

[24] SCHEFFÉ, H. (1943). On a measure problem arising in the theory of nonparametric tests. *Ann. Math. Statist.* **14** 227–233.

[25] TERRY, M. E. (1952). Some rank order tests which are most powerful against specific parametric alternatives. *Ann. Math. Statist.* **23** 346–366.

[26] VAN EEDEN, C. (1963). The relation between Pitman's asymptotic relative efficiency of two tests and the correlation coefficient between their test statistics. *Ann. Math. Statist.* **34** 1442–1451.