The Annals of Statistics
1991, Vol. 19, No. 4, 2254–2258

OVERWEIGHT TAILS ARE INEFFICIENT[1]

BY R. A. LOCKHART

Simon Fraser University

Test statistics which are almost determined by $o(n)$ tail order statistics are shown to provide tests of asymptotic relative efficiency 0 against the usual type of contiguous alternative. The result is applied to several goodness-of-fit tests: the variance weighted Kolmogorov–Smirnov statistic, the Kolmogorov-Smirnov statistic in the stabilized probability plot and the correlation coefficient in a Q–Q plot for a variety of distributions with exponential tails.

Basic results. Suppose $U_1 < \cdots < U_n$ are the order statistics for a sample of size n from a continuous distribution on the unit interval. Goodness-of-fit tests of the hypothesis that this distribution is the uniform distribution, $F(t) = t$, are often based on a comparison of the U's with values predicted by the null hypothesis. Examples include tests based on the empirical process, $n^{1/2}(F_n(t) - F(t))$, where F_n is the empirical distribution of the U's, and tests based on the linearity of a Q–Q plot of the U's.

Tail order statistics, U_i, with i/n near 0 or 1 are less variable than central order statistics with i/n near $1/2$. A number of goodness-of-fit tests compensate for this reduced variability by putting extra weight on the tails. This tactic often produces more powerful tests. For example, in a wide range of Monte Carlo studies, Stephens (1986a) has shown that the Cramér–von Mises test is generally less powerful than the Anderson–Darling test (though there are, of course, many specific alternatives for which this is not true). It is possible to apply too much weight to the tails, however. In this article we establish that any statistic which is essentially determined by too few (smaller order than n) tail order statistics has 0 asymptotic efficiency relative to standard tests against any sequence of contiguous alternatives of the usual type. We show that our results apply to the variance weighted Kolmogorov–Smirnov statistic and to a number of tests based on the correlation coefficient in the Q–Q plot. We begin in the context of the simple null hypothesis given above and then indicate that the result applies rather broadly.

A sequence of events of the form $(U_1, \dots, U_k, U_{n-k}, \dots, U_n) \in C$ is a tail sequence if the sequence $k = k(n)$ satisfies $k/n \to 0$. (Throughout this article quantities designated by Roman letters depend on n and the dependence is suppressed. Quantities designated by Greek letters do not depend on n.) Let G be a sequence of alternatives contiguous to F. Let $r = dG/dF$ be the likelihood ratio, that is, the Radon–Nikodym derivative of G relative to F which

Received September 1989; revised August 1990.
[1]Supported by the Natural Sciences and Engineering Research Council of Canada.
AMS 1980 *subject classifications*. Primary 62G20; secondary 62G30.
Key words and phrases. Goodness of fit, probability plots, correlation tests.

we assume to exist. Set $h = n^{1/2}(r - 1)$. Let P_G denote the distribution of $(U_1, \dots, U_k, U_{n-k}, \dots, U_n)$ under G and similarly for F.

THEOREM 1. *Assume that the sequence h converges to some η in $L_2[0,1]$. If C is a tail sequence, then*

$$|P_G(C) - P_F(C)| \to 0.$$

Suppose $T = T(U_1, \dots, U_n)$ is some sequence of statistics converging in distribution under the null hypothesis to some random variable, say τ.

COROLLARY 1. *Assume that the sequence h converges to some η in $L_2[0,1]$. If there is a sequence of statistics $T^*(U_1, \dots, U_k, U_{n-k}, \dots, U_n)$ with $k = o(n)$ and $|T - T^*| \to 0$ in probability under F, then $T \to \tau$ in distribution under G.*

The proofs rely on the following modification of Le Cam's third lemma; see Hájek and Šidák (1967), page 208ff.

LEMMA. *Suppose Q is a sequence of measures contiguous to P. If T is a sequence of statistics such that* $(\log dQ/dP, T)$ *converges in distribution under P to (Λ, τ) with Λ and τ independent, then T converges in distribution to τ under Q.*

The theorem says that the power of any sequence of tests whose critical regions are a tail sequence will converge to the level of that test. The corollary says that T does not provide a good test of the hypothesis. Typically, such tests as the Kolmogorov–Smirnov, Cramér–von Mises or Anderson–Darling will have limiting power greater than their level along the sequence of alternatives described; this limiting power tends to 1 as the norm of η tends to ∞.

The conclusion may be rephrased to say that the asymptotic efficiency of tests based on T is 0 relative to the likelihood ratio test, or to any other test with nonzero asymptotic efficiency relative to the likelihood test.

REMARK. The hypothesis that $h \to \eta$ in L_2 may be replaced in both the theorem and corollary by the requirement that all but finitely many of the h belong to some fixed compact subset ϕ of L_2. Any counterexample sequence to this apparently stronger conclusion would contain an L_2-convergent subsequence which would provide a counterexample to the theorem or corollary. Similarly, for any fixed sequence $k = k(n)$, the theorem implies $\sup_C |P_G(C) - P_F(C)| \to 0$.

REMARK. The hypothesis that the sequence T converges in distribution may be replaced in the corollary by the requirement that T be tight under F. The conclusion must then be stated in terms of distance between the distributions of T under G and F, where distance is measured by any complete metric for convergence in distribution. T may take values in any separable metric space so that the result applies to functional central limit theorems.

Example. One version of the variance weighted Kolmogorov–Smirnov statistic is $D_W = \sup\{|\hat{F}(u) - u|/(u(1-u))^{1/2}, 0 < u < 1\}$, where $\hat{F}$ is the empirical distribution function of $(U_1, \dots, U_n)$. Jaeschke (1979) shows that $b_n D_W - c_n$ converges in distribution to the standard extreme value distribution $\exp(-\exp(-t))$, where $b_n^2 = 2n \log\log n$ and $c_n = 2\log\log n + 2^{-1}\log\log\log n - 2^{-1}\log(\pi/4)$. Standard weak convergence results show that for each fixed $e > 0$ we have $b_n D_W(e) - c_n \to -\infty$, where $D_W(e) = \sup\{|\hat{F}(u) - u|/(u(1-u))^{1/2}, e < u < 1 - e\}$. It follows that there is a sequence $e = e(n)$ decreasing to 0 for which this convergence holds. For this sequence we find $P(D_W = \sup\{|\hat{F}(u) - u|/(u(1-u))^{1/2}, 0 < u < e \text{ or } 1 - e < u < 1\}) \to 1$. The conditions of the corollary can now be verified.

The result extends to several modifications of D_W; see Shorack and Wellner (1986), pages 597–603. It also extends to a statistic, D_{SP}, due to Michael (1983), namely, the Kolmogorov–Smirnov statistic in the stabilized probability plot. Here

$$\pi D_{\mathrm{SP}} = \max\left\{\left|\arcsin\left(U_i^{1/2}\right) - \arcsin\left(\left(\frac{i}{n+1}\right)^{1/2}\right)\right|; i\ 1 \le i \le n\right\};$$

see Lockhart and Stephens (1991) for a discussion of tests based on these plots.

Composite null hypotheses. The theorem applies to the case of composite null hypotheses. Suppose the data are $Y_1 < \cdots < Y_n$, the order statistics for a sample of size n from a continuous distribution. Suppose that the null hypothesis is a family of distributions $\mathscr{F}$. We want to calculate the approximate power along a sequence of alternatives G contiguous to a sequence F of members of $\mathscr{F}$. Define h as above. Since the order statistics Y_i may be constructed as $Y_i = F^{-1}(U_i)$, the theorem and corollary may be applied. Note that the conditions on h in the theorem and corollary become conditions on $h \circ F^{-1}$ here. Often F is some fixed member of $\mathscr{F}$ and the condition that $h \circ F^{-1}$ converge in $L_2[0,1]$ becomes the condition that h converge in $L_2(F)$.

Example. When $\mathscr{F}$ is the location-scale family $\{F(x) = \Phi((x - \alpha)/\beta), \alpha \in \mathbb{R}, \beta > 0\}$ tests may be based on the linearity of a Q–Q plot, that is, a plot of Y_i against a measure of the centre of the distribution of Y_i under the standard member of $\mathscr{F}$ such as $\Phi^{-1}(i/(n+1))$; see Stephens (1986b) for an extensive discussion. One natural statistic is the correlation coefficient in this plot. When Φ is the standard normal cumulative, the resulting test is very powerful; it is asymptotically equivalent to the Shapiro–Wilk test [see Leslie, Stephens and Fotopolous (1986)].

For distributions with somewhat heavier tails such as the exponential, extreme value and logistic distributions, the situation is dramatically different. McLaren and Lockhart (1986) noted that the correlation coefficient satisfies the hypotheses of the theorem and used their observation to show that the correlation coefficient has zero ARE for a variety of contiguous alternatives.

The present result simplifies their Proposition 2 and may be applied in every case they studied. Again the conclusion is that the power and level of the test have the same limit and the ARE of the test is 0.

PROOF OF THEOREM 1. We compute under the null hypothesis. There is no loss in assuming that $P_F(C)$ is a convergent sequence. The log-likelihood ratio statistic is $L = \sum_i \log(1 + h(U_i)/n^{1/2})$. Uniform order statistics can be constructed as $U_i = (V_1 + \cdots + V_i)/(V_1 + \cdots + V_{n+1})$, where $V_1, \dots, V_{n+1}$ are independent standard exponentials. Guttorp and Lockhart (1988), pages 435 and 444, show that

$$L = \sum_{1}^{n+1} h_i(V_i - 1) - \int \eta^2/2 + o_P(1),$$

where $h_i = (n+1)^{1/2} \int h(x) 1(i - 1 < (n+1)x < i)\, dx$. Implicit in their argument is the fact that $\sum_{[na]}^{[nb]} h_i^2 \to \int_a^b \eta^2(x)\, dx$ for each $0 \le a \le b \le 1$. A variance calculation then shows that

$$L = \sum_{k+1}^{n-k} h_i(V_i - 1) - \int \eta^2/2 + o_P(1).$$

Now $(U_i - U_{i-1})/(U_{n-k} - U_k) = V_i/(V_{k+1} + \cdots + V_{n-k})$. Since $(V_{k+1} + \cdots + V_{n-k})/n \to 1$ in probability and $\sum h_i = \int_0^1 h = 0$, we have

$$L = n^{-1} \sum_{k+1}^{n-k} h_i \frac{U_i - U_{i-1}}{U_{n-k} - U_k} - \int \eta^2/2 + o_P(1) \equiv L_1 + o_P(1).$$

The conditional distribution of $(U_{k+1}, \dots, U_{n-k-1})$ given $(U_1, \dots, U_k, U_{n-k}, \dots, U_n)$ is that of order statistics for a sample of size $n - 2k - 1$ from the uniform distribution on the interval (U_k, U_{n-k}). Letting $W_i = (U_{k+i} - U_k)/(U_{n-k} - U_k)$ for $i = 1, \dots, n - 2k - 1$, we see that the conditional law of $W_1, \dots, W_{n-2k-1}$ is that of the order statistics for a sample of size $n - 2k - 1$ from the uniform distribution on the unit interval. Since this is free of $(U_1, \dots, U_k, U_{n-k}, \dots, U_n)$ and since L_1 is a function of $W_1, \dots, W_{n-2k-1}$, we see that L_1 is independent of C. Hence $(1_C, L)$ converge jointly in distribution to a distribution with independent marginals, the marginal limiting distribution of L being $N(-\sigma^2/2, \sigma^2)$, where $\sigma^2 = \int \eta^2(x)\, dx$. The theorem now follows from the lemma. □

PROOF OF THE LEMMA. Fix x a continuity point of τ. Let R be the probability $(P + Q)/2$ which dominates P and Q. We have

$$\begin{aligned} Q(T \le x) &= E_R\left(\frac{dQ}{dR} 1(T \le x)\right) \\ &= E_R\left(\frac{dQ}{dP}\frac{dP}{dR} 1(T \le x)\right) + Q\left(\frac{dP}{dR} = 0, T \le x\right) \\ &= E_P\left(\frac{dQ}{dP} 1(T \le x)\right) + Q\left(\frac{dP}{dR} = 0, T \le x\right). \end{aligned}$$

The assumed contiguity shows that the second term converges to 0 since $P(dP/dR = 0) = 0$. Joint convergence in distribution shows that for each fixed c, a continuity point of the distribution of $|\Lambda|$, we have

$$E_P(\exp(\log dQ/dP)1(T \le x)1(-c \le \log dQ/dP \le c)) - \text{Prob}(\tau \le x)E(\exp(\Lambda)1(-c \le \Lambda \le c)) \to 0.$$

There is then a sequence c, of continuity points of Λ, converging to ∞ so slowly that this difference still converges to 0. For this sequence c we see that

$$E_P(\exp(\log dQ/dP)1(T \le x)1(|\log dQ/dP| > c)) \le Q(|\log dQ/dP| > c).$$

On the one hand, $Q(\log dQ/dP < -c) \le \exp(-c) \to 0$. On the other hand,

$$P(\log dQ/dP > c) = P(\log dP/dQ < -c) \le \exp(-c) \to 0$$

so that by contiguity $Q(|\log dQ/dP| > c) \to 0$. Recall that contiguity guarantees $E(\exp(\Lambda)) = 1$ to finish the proof. □

Acknowledgments. I am grateful to an Associate Editor and referees for several helpful comments which improved the presentation and strengthened the results.

REFERENCES

GUTTORP, P. and LOCKHART, R. A. (1988). On the asymptotic distribution of quadratic forms in uniform order statistics. *Ann. Statist.* **16** 433–449.

HÁJEK, J. and ŠIDÁK, Z. (1967). *Theory of Rank Tests*. Academic, New York.

JAESCHKE, D. (1979). The asymptotic distribution of the supremum of the standardized empirical distribution function on subintervals. *Ann. Statist.* **7** 108–115.

LESLIE, J. R., STEPHENS, M. A. and FOTOPOLOUS, S. (1986). Asymptotic distribution of the Shapiro–Wilk W test for normality. *Ann. Statist.* **14** 1497–1506.

LOCKHART, R. A. and STEPHENS, M. A. (1991). On EDF tests and tests based on probability plots: Some connections. Technical Report, Dept. Mathematics and Statistics, Simon Fraser University.

MCLAREN, C. G. and LOCKHART, R. A. (1986). On the asymptotic efficiency of certain correlation tests of fit. *Canad. J. Statist.* **15** 159–167.

MICHAEL, J. R. (1983). The stabilized probability plot. *Biometrika* **70** 11–17.

SHORACK, G. R. and WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.

STEPHENS, M. A. (1986a). Tests based on EDF statistics. In *Goodness-of-Fit Techniques* (R. B. D'Agostino and M. A. Stephens, eds.) 97–193. North-Holland, Amsterdam.

STEPHENS, M. A. (1986b). Tests based on regression and correlation. In *Goodness-of-Fit Techniques* (R. B. D'Agostino and M. A. Stephens, eds.) 195–233. North-Holland, Amsterdam.

DEPARTMENT OF MATHEMATICS
AND STATISTICS
SIMON FRASER UNIVERSITY
BURNABY, BRITISH COLUMBIA
CANADA V5A 1S6