# ON LARGE SAMPLE PROPERTIES OF CERTAIN NONPARAMETRIC PROCEDURES

HERMAN RUBIN
PURDUE UNIVERSITY

## 1. Summary and introduction

Efficiencies of one sided and two sided procedures are considered from the standpoint of risk. It is shown that the two sided Kolmogorov-Smirnov (K-S) and Kuiper procedures, which were shown in [4] to be asymptotically equiefficient with the median for translation alternatives for symmetric unimodal distributions, have efficiencies for sample sizes in a wide range in the general vicinity of that of the median; but even if certain standard asymptotic approximations can be made, the efficiencies are not too close to that of the median, and in many cases the dominant asymptotic correction term does not even yield the sign of the deviation for samples of size $10^{20}$.

A procedure briefly discussed in [1], for which the Pitman efficiency is zero, has good Bayes risk efficiency for translation alternatives for any distribution and merits further work for two sided testing.

In the one sided case, the one sided K-S procedure appears to be somewhat worse to much worse than a procedure introduced by the author in [3]. Also, the K-S procedure involves a choice of significance level which is highly distribution dependent.

We shall consider the "moderately large sample" efficiencies of certain well known and not sufficiently well known nonparametric procedures from a decision theoretic standpoint. By "moderately large sample" we shall mean that central limit type theorems yield adequate approximations to the distributions involved, but that the further asymptotic approximations of the type in [4] are not necessarily very good. We shall also assume that the samples under consideration are sufficiently large that the large sample form of the risk can be used.

That is, we shall carry out our computations as if the observations can be considered as a stochastic process on [0, 1] such that

$$X(t) = \theta h(t) + Y(t), \tag{1.1}$$

Research was supported in part by the Office of Naval Research Contract N00014-67-A-226-0014, project number NR042-216 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

where $Y$ is a separable Gaussian process with mean 0 and covariance function

$$\sum (t, u) = 4[\min (t, u) - tu], \tag{1.2}$$

and $h(t)$ is a multiple of $fF^{-1}(t)$, chosen so that $h(0.5) = 1$. The choice of these normalizing factors is for computational convenience; the median corresponds to $-X(0.5)$ in standard units. For simulation purposes, we have chosen five distributions:

normal, with

$$h_N(t) = \exp \left\{-0.5 \{[N(0, 1)(t)]^{-1}\}^2\right\}; \tag{1.3}$$

logistic, with

$$h_L(t) = 4t(1 - t); \tag{1.4}$$

double exponential, with

$$h_D(t) = \begin{cases} 2t, & t \leqq 0.5, \\ 2(1 - t), & t > 0.5; \end{cases} \tag{1.5}$$

Cauchy, with

$$h_C(t) = \sin^2 \pi t; \tag{1.6}$$

and a distribution with density $C(1 + \tau|x - \theta|)^{-10/9}$, in which case

$$h_T(t) = \begin{cases} (2t)^{10}, & t \leqq 0.5, \\ [2(1 - t)]^{10}, & t > 0.5. \end{cases} \tag{1.7}$$

The loss structure was taken to be $2|\theta|d\theta$ for a wrong decision in the one sided testing problem [2]—it can be strongly argued that for "reasonably large" samples no other loss function is reasonable for this problem. For the two sided problem the weight function was taken, as in [4], to be 1 if a type I error is made, and $|\theta|^k \, d\theta/\sqrt{2\pi} \, \mu_k$ for a type II error, where $\mu_k$ is the $k$th absolute moment of the normal distribution. The choice of multiplicative constants was chosen so that if $Z$ is $N(\theta, \sigma^2)$, then it will never pay to accept the null hypothesis if $\sigma > 1$, but for $Z$ sufficiently small it will pay if $\sigma < 1$.

In the two sided case these normalizations correspond to establishing a base for the sample size. In the one sided case, if the value of the translation parameter at which there is indifference is $\theta^*$, the risk is $E(\theta^{*2})$.

The procedures we have evaluated by Monte Carlo are, for the two sided problem, Kolmogorov-Smirnov and Kuiper; and we have compared them to the median, for which it is known [4] that they are asymptotically equiefficient. For the one sided case, the Kolmogorov-Smirnov statistic has been compared with a symmetrized version introduced by the author in [3].

## 2. Two sided tests: asymptotic treatment

The procedures that we shall consider are the median, Kolmogorov-Smirnov, and Kuiper. We shall also consider a test, suggested in [2], for which the Pitman efficiency is 0, but whose Bayes risk efficiency is that of the best order statistic. For the median (in our approximation: $X(0.5) + \theta h(0.5)$) the probability of exceeding $C$ under the null hypothesis is

$$P_M = \frac{2}{\sqrt{2\pi}} \int_c^{\infty} e^{-t^2/2}\, dt \sim \frac{2}{\sqrt{2\pi}\, c} e^{-c^2/2}. \tag{2.1}$$

For the K-S statistic ($\sup |X(t) + \theta h(t)|$) the corresponding probability is

$$P_{K\text{-}S} = 2\Sigma(-1)^{n-1} e^{-c^2 n^2/2} \sim 2e^{-c^2/2}, \tag{2.2}$$

and for the Kuiper statistic, the probability is

$$P_K = 2\Sigma(n^2 c^2 - 1) e^{-c^2 n^2/2} \sim 2c^2 e^{-c^2/2}. \tag{2.3}$$

(Note that there is a scale factor of 2 in the expressions for $P_K$ and $P_{K\text{-}S}$).

Now let us examine what happens under the alternative. Let $X^+(\theta) = \sup (X(t) + \theta h(t))$. For $\theta$ reasonably large, if $h(t) \sim 1 - \lambda|t - \frac{1}{2}|^{\gamma}$, $\gamma > \frac{1}{2}$, $X^+(\theta)$ is approximately $\theta + \lambda^{-1/\gamma}\theta^{-1/\gamma} Y_\gamma - Z$, where $Z$ is normal $(0, 1)$ and $Y_\gamma$ is a positive random variable whose distribution is not known except for $\gamma = 1$. Hence that $\theta$ for which $X^+(\theta) = c$ is approximately

$$\theta_c = c + Z - \lambda^{-1/\gamma} c^{-1/\gamma} Y_\gamma. \tag{2.4}$$

Therefore

$$E_{K\text{-}S}(\theta_c^k) \sim c^k + \binom{k}{2} c^{k-2} - kK_\gamma c^{k-1-1/\gamma}. \tag{2.5}$$

For the Kuiper statistic we also need $X^-(\theta) = \inf (X(t) + \theta h(t))$. Here if $\theta$ is reasonably large and $h(t) \sim pt^\beta$, $\beta > \frac{1}{2}$, $X^-(\theta) \sim -p^{-1/\beta}\theta^{-1/\beta} W_\beta$, and

$$E_K(\theta_c^k) \sim c^k + \binom{k}{2} c^{k-2} - kK_\gamma c^{k-1-1/\gamma} - kH_\beta c^{k-1-1/\beta}. \tag{2.6}$$

For the distributions we are considering, the values of $\gamma$ and $\beta$ are shown in Table I. (For the normal, the tail behavior is slightly more complicated, but

TABLE I

Values of $\gamma$ and $\beta$

| | $\gamma$ | $\beta$ |
|---|---|---|
| normal | 2 | 1 |
| logistic | 2 | 1 |
| double exponential | 1 | 1 |
| Cauchy | 2 | 2 |
| long tailed | 1 | 10 |

since for the Kuiper statistic the larger of $\beta$ and $\gamma$ is what counts, this is not a problem.)

Incidentally, in the case of the median,

$$E_M(\theta_c^k) \sim c^k + \binom{k}{2} c^{k-2}. \tag{2.7}$$

Note that for the K-S and the Kuiper statistic, $E(\theta_c^k)$ is smaller than for the median. However, the $c$ required to obtain a given type I error is somewhat larger.

Now let us investigate what happens for $m$th power loss for samples of size $n$ if the cut off point is $c$. We obtain for the type II risk

$$R_2 = \frac{2}{\sqrt{2\pi}\,\mu_m n^{(m+1)/2}} E\left(\frac{\theta_c^{m+1}}{m+1}\right). \tag{2.8}$$

Hence our combined risk is

$$R = P_I(c) + \left(\frac{c^{m+1}}{m+1} + \frac{m}{2} c^{m-1} - Rc^{m-1/\gamma} + \cdots\right) Bn^{-(m+1)/2}, \tag{2.9}$$

where $P_I(c) \sim Ac + e^{-c^2/2}$. Now a lengthy calculation shows that the dominant correction term to the asymptotic expression

$$R \sim B(m+1)^{(m-1)/2}\left(\frac{\log n}{n}\right)^{(m+1)/2} \tag{2.10}$$

has the relative value

$$C_1 = \frac{1}{2} \frac{(q+1-m)\log\log n}{[(m+1)\log n]^{1/2}}, \tag{2.11}$$

which of course increases with $q$. Thus, for extremely large $n$, the median is better than the K-S test, which is better than the Kuiper test.

However, extremely large depends on $\log\log n$. Since $\log\log 10^{20} < 4$, for practical purposes the next term (which depends on $A$) comes into effect, and the $-Rc^{m-1/\gamma}$ term may actually be dominant.

## 3. Two sided tests: moderately large sample and empirical results

A computation based on the likelihood ratio shows that for small $n$ the K-S and Kuiper statistics are approximately equivalent to the best procedure. (This requires the probability of type I error to be nearly 1.) Apparently this efficiency drops off rapidly. Let us look at the results of Monte Carlo computations (Table II).

The values are independent for the different distributions, but dependent within any one distribution. The standard deviations (estimated from 1000 sample processes) of these efficiencies are 1 to 2 per cent for samples of size 10

TABLE II

EFFICIENCY (PER CENT)

| | | Kolmogorov-Smirnov constant loss | Kolmogorov-Smirnov absolute error | Kolmogorov-Smirnov squared error | Kuiper constant loss | Kuiper absolute error | Kuiper squared error |
|---|---|---|---|---|---|---|---|
| Normal | 1 | 127 | 123 | 123 | — | — | — |
| | 2 | 129 | 126 | 119 | 70 | 72 | 72 |
| | 5 | 122 | 118 | 119 | 65 | 68 | 71 |
| | 10 | 119 | 119 | 118 | 68 | 70 | 73 |
| | $10^2$ | 115 | 116 | 114 | 74 | 76 | 80 |
| | $10^3$ | 117 | 114 | 113 | 75 | 81 | 85 |
| | $10^5$ | 114 | 112 | 110 | 80 | 87 | 91 |
| | $10^{10}$ | 112 | 109 | 107 | 88 | 93 | 96 |
| | $10^{20}$ | 108 | 106 | 105 | 93 | 97 | 98 |
| Logistic | 1 | 111 | 114 | 116 | — | — | — |
| | 2 | 117 | 117 | 111 | 73 | 75 | 76 |
| | 5 | 114 | 112 | 113 | 69 | 72 | 74 |
| | 10 | 111 | 113 | 113 | 72 | 73 | 75 |
| | $10^2$ | 110 | 112 | 111 | 75 | 78 | 81 |
| | $10^3$ | 111 | 111 | 110 | 77 | 82 | 86 |
| | $10^5$ | 110 | 109 | 108 | 82 | 88 | 91 |
| | $10^{10}$ | 109 | 107 | 106 | 89 | 93 | 96 |
| | $10^{20}$ | 107 | 105 | 104 | 94 | 97 | 98 |
| Double exponential | 1 | — | — | — | — | — | — |
| | 2 | 79 | 79 | 80 | 64 | 63 | 67 |
| | 5 | 78 | 80 | 82 | 60 | 61 | 62 |
| | 10 | 79 | 82 | 83 | 61 | 63 | 66 |
| | $10^2$ | 82 | 86 | 87 | 64 | 68 | 72 |
| | $10^3$ | 85 | 88 | 90 | 67 | 72 | 76 |
| | $10^5$ | 88 | 91 | 93 | 72 | 78 | 82 |
| | $10^{10}$ | 92 | 94 | 96 | 79 | 85 | 88 |
| | $10^{20}$ | 94 | 96 | 97 | 86 | 90 | 92 |
| Cauchy | 1 | — | — | — | — | — | — |
| | 2 | 87 | 90 | 92 | 85 | 87 | 89 |
| | 5 | 88 | 91 | 91 | 80 | 84 | 87 |
| | 10 | 89 | 93 | 94 | 81 | 85 | 87 |
| | $10^2$ | 93 | 97 | 98 | 84 | 89 | 92 |
| | $10^3$ | 97 | 99 | 100 | 89 | 93 | 95 |
| | $10^5$ | 98 | 101 | 101 | 92 | 96 | 98 |
| | $10^{10}$ | 103 | 102 | 102 | 97 | 99 | 101 |
| | $10^{20}$ | 102 | 102 | 102 | 99 | 101 | 102 |
| Long tailed | 1 | — | — | — | 100+ | 105 | 112 |
| | 2 | 70 | 73 | 77 | 90 | 96 | 102 |
| | 5 | 66 | 65 | 65 | 78 | 86 | 91 |
| | 10 | 57 | 61 | 64 | 81 | 86 | 94 |
| | $10^2$ | 61 | 68 | 73 | 81 | 93 | 100 |
| | $10^3$ | 66 | 74 | 79 | 88 | 98 | 104 |
| | $10^5$ | 74 | 82 | 86 | 94 | 103 | 107 |
| | $10^{10}$ | 83 | 89 | 92 | 103 | 107 | 109 |
| | $10^{20}$ | 93 | 94 | 95 | 106 | 108 | 108 |

and, with very few exceptions, 0.1 to 0.2 per cent for samples of size $10^{20}$. Thus, while individual figures for small sample sizes are not too reliable, the general picture is clear: for the Kolmogorov-Smirnov test, the flatness at the median determines the efficiency, and for samples of size $10^{20}$ relative to the base, the dominant asymptotic error has yet to make its presence felt.

The results are also similar for the Kuiper statistic. Several cases also clearly show the dip for small samples in the efficiencies. These results also agree with the exact calculations for K-S with 0th power loss for the double exponential in [3]. The optimal significance levels also are not much affected by the test.

A test occasionally considered (see, for example [1]) is to use $T_n = \sqrt{n}\sup|(F_n - F)/[F(1 - F)]^{1/2}|$. The statistic $T_n$ is more sensitive to deviations in the tails than the K-S statistic. Now examination of

$$T_n(x) = \sqrt{n}\,\frac{F_n(x) - F(x)}{[F(x)(1 - F(x))]^{1/2}} \tag{3.1}$$

by the usual methods shows that

$$T_n \sim (2 \log \log n)^{1/2}. \tag{3.2}$$

A further examination shows that the statistic cannot be very sensitive to Pitman alternatives since that $x$ for which $T_n = |T_n(x)|$ is likely to be near 0 to 1. Of course, $(2 \log \log 10^{20})^{1/2} < 2.15\,(2 \log \log 10)^{1/2}$, so that even this argument may not be too serious for reasonable sample sizes. But we note that for $k$th power loss, for the K-S test the critical deviation is approximately $\frac{1}{2}[(k + 1) \log n]^{1/2}$, which grows much more rapidly. However, if we break the ordered observations below the median into groups of size 1, 2, 4, 8, $\cdots$, and examine the distribution of $T_n(x)$ in the corresponding intervals, we find that

$$P(T_n > c) < K \log (n + 1)e^{-c^2/2}. \tag{3.3}$$

This shows that from the Bayes risk standpoint, this statistic bears much the same relationship to the best order statistic as the K-S or Kuiper statistic does to the median! This test consequently merits investigation.

## 4. One sided tests

If the weight function is $2|\theta|\,d\theta$ for a wrong decision, and if the structural model is such that the observation $Y = \phi(\theta, X)$, and for $\theta < \hat{\theta}(x)$ the decision is made that $\theta < 0$, the risk is $E[(\hat{\theta}(x))^2]$. This calculation can be applied in our model to the one sided K-S tests and also to the symmetric test given by the author in [3]. The symmetric test has similar properties to the median for all symmetric unimodal distributions; its reciprocal efficiency relative to the median is between $2 - \frac{1}{6}\pi^2 = 0.355$ and a number bounded by $(\frac{1}{3}\sqrt{\pi} + 1)^2 \sim 7.9$.

The one sided K-S test does not fare so well. From equation (1.1), note that if $\theta$ is large the maximum of $X(t)$ will be large with large probability since $X(\frac{1}{2}) = \theta + Y(\frac{1}{2})$, but the minimum may still be quite negative if $\theta$ is not very large, especially if $h$ is small some distance away from 0. This is indeed borne out by the empirical results. Unfortunately, it was not anticipated just how bad things would get, and hence it is necessary to crudely estimate some numbers. The results are given in Table III with standard errors in parentheses.

TABLE III

VARIANCE OF INDIFFERENCE POINT

| | Normal | Logistic | Double exponential | Cauchy | Long tail |
|---|---|---|---|---|---|
| Symmetric | .735(.033) | .796(.034) | 1.405(.061) | 1.386(.062) | 6.13(.19) |
| K-S, one sided | .851(.027) | .942(.029 | 1.737(.057) | 1.902(.073) | ~11(~3) |
| K-S, one sided 50 per cent | .852(.017) | .950(.029) | 1.931(.07) | ~4.5 | ~160 • |
| Optimal level | .49 | .47 | .39 | .28 | .008 |

Again, the values for the one sided K-S at optimal level for the double exponential and the optimal level agree very well with the theoretical values of $\sqrt{14} - 2 = 1.74166$ and $e^{\sqrt{0.875}} = 0.39244$, respectively. Note that for not too bad distributions, the one sided K-S test is fairly good if used at the optimal level, which varies considerably with the distribution. If the 50 per cent level were used, as is optimal for any symmetric test statistic, the tail of the Cauchy is already bad enough to cause problems. It was not anticipated in the empirical procedure that when $\theta$ was chosen to make the maximum of $X(\theta)$ greater than 12.8 (or the minimum less than $-12.8$), which is beyond the $10^{-33}$ level for the Kuiper statistic that there would be any significant problems with the minimum (maximum). A few values for the Cauchy distribution were far enough out to give questionable accuracy at the 50th percentile; for the long tailed distribution the figures given are probably slightly conservative.

◇ ◇ ◇ ◇ ◇

The author wishes to express his thanks to Glennis Cohen for her assistance in producing the programs for the simulation and processing of the numerical results, and to Arthur Rubin for his invaluable help in the debugging of these programs.

REFERENCES

[1] T. W. ANDERSON and D. A. DARLING, "Asymptotic theory of certain 'goodness of fit' criteria based on stochastic processes," *Ann. Math. Statist.*, Vol. 23 (1952) pp. 193–212.

[2] H. CHERNOFF, "Sequential tests for the mean of a normal distribution," *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, 1961, Vol. 1, pp. 79–91.

[3] H. RUBIN, "Decision-theoretic evaluation of some non-parametric methods," *Nonparametric Techniques in Statistical Inference* (edited by M. L. Puri, London), Cambridge University Press, 1970, pp. 579–583.

[4] H. RUBIN and J. SETHURAMAN, "Bayes risk efficiency," *Sankhyā Ser. A*, Vol. 27 (1965) pp. 347–356.