Plotting with confidence. Graphical comparisons of two populations.

By Kjell A. Doksum
University of California, Berkeley
and
Gerald L. Sievers
Western Michigan University, Kalamazoo.

SUMMARY

Statistical methods that give detailed descriptions of how populations differ are considered. These descriptions are in term of a response function $\Delta(x)$ with the property that $X + \Delta(X)$ has the same distribution as Y. The methods are based on simultaneous confidence bands for the response function computed from independent samples from the two populations. Both general and parametric models are considered and comparisons between the various methods are made.

Some key words : Response function; Confidence bands; Shift function; Behrens-Fisher model; Q-Q plots; Nonlinear models; Empirical probability plot; Two sample problem.

1. INTRODUCTION

We consider the problem of comparing two populations with distribution functions F and G on the basis of two independent random samples $X_1,\ldots,X_m$ and $Y_1,\ldots,Y_n$, respectively. Instead of the usual shift model where $F(x) = G(x+\theta)$ for all x , we treat the general case where $F(x) = G(x+\Delta(x))$ for some function $\Delta(x)$. If the X's are control responses and the Y's are treatment responses, $\Delta(x)$ can under certain conditions be regarded as the amount the treatment adds to a potential control response x, (Doksum (1974)). Since it gives the effect of the treatment as a function of the response variable, we call it the <u>response function.</u> Under general conditions it is the only function of x that satisfies $X + \Delta(X) \sim Y$, where $\sim$ denotes distributed as. Thus $\Delta(\cdot)$ is the amount of "shift" needed to bring the X's up to the Y's in distribution and it is also referred to as the <u>shift function.</u>

Assume that F and G are continuous. Let F^{-1} denote the left inverse of F . Then we can write

$$\Delta(x) = G^{-1}(F(x)) - x .$$

If in fact a shift model holds, that is, $F(x) = G(x+\theta)$ for some constant θ , then $\Delta(x) \equiv \theta$.

A natural estimate of $G^{-1}(F(x))$ is $G_n^{-1}(F_m(x))$, where F_m and G_n denote the empirical distribution functions based on the X and Y samples. The Q-Q plot considered by Wilk and Gnanadesikan (1968) is essentially $G_n^{-1}F_m$ evaluated at the X order statistics. Doksum (1974) referred to it as the empirical probability plot and derived the asymptotic distribution of $\hat{\Delta}(x) = G_n^{-1}(F_m(x)) - x$.

Suppose that a beneficial treatment leads to large responses. Then certain natural questions arise : (i) Is the treatment beneficial for all the members of the population, i.e., is $\Delta(x) > 0$ for all x ? (ii) If not, for which part of the population is the treatment beneficial, i.e., what is $\{x:\Delta(x) > 0\}$?

The kind of model that is in effect also yields information about how the treatment works and about which statistical analysis is appropriate. Thus the following questions are of interest : (iii) Does a shift model hold, i.e.,is $\Delta(x) \equiv \theta$, some θ , all x ? (iv) If not, does a shift-scale model hold, i.e., is $\Delta(x) = \alpha + \beta x$, some α , β , all x ?

These questions can be answered by giving a simultaneous (in x) confidence band $[\underline{\Delta}(x),\bar{\Delta}(x)]$ for $\Delta(x)$. Thus (i) is answered in the affirmative if $\underline{\Delta}(x) > 0$ for all x , (ii) has solution $\{x:\underline{\Delta}(x) > 0\}$, (iii) is rejected if no horizontal line fits in the confidence band, and (iv) has a negative response if no straight line fits in the band. Note that the first two answers only required a lower confidence boundary $\underline{\Delta}(x)$.

Such confidence bands have been considered by Doksum (1964), Switzer (1975), and Sievers (1975). The latter two papers derive a band based on the two sample Kolmogorov-Smirnov statistic and thereby obtain the exact confidence coefficient of the band in the former paper. Similar bands have been considered by Steck, Zimmer and Williams (1972) in connection with the "acceleration function" GF^{-1} . Here we consider bands based on statistics of the form

$$\sup_{a_m \leq x \leq b_m} |F_m(x)-G_n(x)|/\varphi(H_N(x))$$

where

$$H_N(x) = \lambda F_m(x) + (1-\lambda)G_n(x) , \quad \lambda=m/N$$

Efficiency comparisons are made between such bands in terms of square ratios of widths, and it is found that the choice $\varphi(u) = \sqrt{u(1-u)}$ leads to an efficient band.

In the case that a location-scale model can be assumed, a band which improves on the above general bands is constructed from order statistics and its asymptotic efficiencies with respect to the general bands are given.

For the normal Behrens-Fisher model, the likelihood ratio band is derived and it is shown to be much more efficient than the general bands in the normal model.

2. NONPARAMETRIC SIMULTANEOUS CONFIDENCE BANDS FOR THE RESPONSE FUNCTION.

In the nonparametric case, it is natural to construct confidence bands for $\Delta(x)$ using pivots based on empirical distribution functions F_m and G_n. The key to finding such pivots is to note that $G_{\Delta,n}$ defined by

$$G_{\Delta,n}(y) = G_n(\Delta(y) + y)$$

is distributed as the empirical distribution of a sample of size n from F. To see this note that

$$G_n(\Delta(y) + y) = [\text{No. of } Y_i \leq G^{-1}(F(y))]/n$$
$$= [\text{No. of } F^{-1}(G(Y_i)) \leq y]/n$$

and $F^{-1}(G(Y))$ has distribution F.

Now if $\emptyset(F_m, G_n)$ is a distribution-free level α test function for $H_0: F = G$,
then

$$\{\Delta(.) : \emptyset(F_m, G_{\Delta,n}) = 0\} \tag{1}$$

is a distribution-free, level $(1 - \alpha)$ confidence region for the response function $\Delta(\cdot)$.

These regions will reduce to simple bands if we consider distribution-free test statistics $T(F_m, G_n)$ with the property that the inequality $T(F_m, G_n) \leq K$ is equivalent to

$$\underline{h}(F_m(x)) \leq G_n(x) \leq \overline{h}(F_m(x)) \quad \text{for all } x \tag{2}$$

for some functions $\underline{h}$ and $\overline{h}$. Typically these functions are nondecreasing. For instance, let $N = m + n$, $M = mn/N$ and suppose

$$T(F_m, G_n) = D_N = \sqrt{M} \sup_x |F_m(x) - G_n(x)| ,$$

the Kolmogorov-Smirnov statistic. Then $\underline{h}(x) = x - K/\sqrt{M}$ and $\overline{h}(x) = x + K/\sqrt{M}$.

From (2) , we derive confidence bands as follows. Let $G_n^{-1}(u) = \inf\{x : G_n(x) \geq u\}$ and $G_n^{-I}(u) = \sup\{x : G_n(x) \leq u\}$ be the left and right inverses of G_n , and suppose that K is chosen so that

$$P_{F=G}(T(F_m, G_n) \leq K) = 1 - \alpha . \tag{3}$$

Then

$$\begin{aligned} 1 - \alpha &= P_{F=G}(T(F_m, G_n) \leq K) \\ &= P_{F,G}(T(F_m, G_{\Delta,n}) \leq K) \\ &= P_{F,G}(\underline{h}(F_m(x)) \leq G_{\Delta,n}(x) \leq \overline{h}(F_m(x)) \text{ ; all } x) \\ &= P_{F,G}(\underline{h}(F_m(x)) \leq G_n(\Delta(x) + x) \leq \overline{h}(F_m(x)) \text{ ; all } x) \\ &= P_{F,G}(G_n^{-1}(\underline{h}(F_m(x))) - x \leq \Delta(x) < G_n^{-I}(\overline{h}(F_m(x))) - x, \\ &\qquad \text{all } x) . \end{aligned}$$

We have shown that

<u>Proposition 1:</u>

If (2) and (3) hold, then

$$[G_n^{-1}(\underline{h}(F_m(x))) - x , G_n^{-I}(\overline{h}(F_m(x))) - x) \tag{4}$$

as x ranges from $-\infty$ to ∞ gives a level $(1 - \alpha)$, simultaneous, distribution-free confidence band for the response function $\Delta(x)$.

Now suppose that $K_{S,\alpha}$ has been chosen from the Kolmogorov-Smirnov tables (e.g., Kim and Jennrich (1968), (1970)) so that

$$P_{F=G}(D_N \leq K_{S,\alpha}) = 1 - \alpha .$$

Then

Remark 1:

A level $(1 - \alpha)$, simultaneous, distribution-free confidence band for $\Delta(x)$, $-\infty < x < \infty$ is given by

$$\left[G_n^{-1}\left(F_m(x) - \frac{K_{S,\alpha}}{\sqrt{M}}\right) - x \quad , \quad G_n^{-1}\left(F_m(x) + \frac{K_{S,\alpha}}{\sqrt{M}}\right) - x\right) .$$

This band, which we call the S band and denote by $[\underline{\Delta}_S(x) , \overline{\Delta}_S(x))$, was obtained by Switzer (1975) and Sievers (1975) and should replace a similar band given by Doksum (1974).

Let $[t]$ denote the greatest integer less than or equal to t , let $<t>$ be the least integer greater than or equal to t , let $X(1) <\ldots< X(m)$; $Y(1) <\ldots< Y(n)$ denote the order statistics of the X and Y samples, and define $Y(j) = -\infty$ for $j < 0$ and $Y(j) = \infty$ for $j \geq n + 1$. Then the band (4) can be expressed as

$$[\underline{\Delta}(x),\overline{\Delta}(x)) = [Y(<n\underline{h}(\tfrac{i}{m})>) - x, \; Y([n\overline{h}(\tfrac{i}{m})]+1) - x) \qquad (5)$$

for $x \in [X(i), X(i+1))$, $i=0,1,\ldots,m$, with $X(0) = -\infty$ and $X(m+1) = \infty$. This representation was used to produce Fig. 1 where the S band (solid lines)

$$[\underline{\Delta}_S(x), \overline{\Delta}_S(x)) = [Y(<n(\tfrac{i}{m} - \frac{K_{S,\alpha}}{\sqrt{M}})>) - x, Y([n(\tfrac{i}{m} + \frac{K_{S,\alpha}}{\sqrt{M}})] + 1)-x)$$

is given for X and Y samples from $N(0,1)$ and $N(1,4)$ distributions, respectively. In this figure, $m=n=100$ and $\alpha = .05$.

The general method can also be applied to a weighted sup norm statistic

$$W_N = W_N(F_m, G_n) = \sqrt{M} \sup_{\{x: a \leq F_m(x) \leq b\}} \frac{|F_m(x) - G_n(x)|}{\Psi(H_N(x))} \tag{6}$$

where $H_N(x) = \lambda F_m(x) + (1-\lambda)G_n(x)$; $\lambda = m/N$ and $0 \leq a < b \leq 1$. If we choose $\Psi(t) = \sqrt{t(1-t)}$, then we give approximately equal weight to each x in the sense that $\sqrt{M}\,[F_m(x)-G_n(x)]/\Psi(H_N(x))$ has asymptotic variance independent of x . From Borokov and Sycheva (1968) it follows that if we consider one sided (without the absolute value) test statistics in the class (6) with $0 < a < b < 1$ this choice of Ψ asymptotically maximizes the minimum power when testing $H_0: F = G$ vs. $H_1: F(x) - G(x) \geq \delta$ for some x , $\delta > 0$.

To apply Proposition 1 , we need to solve the inequality

$$|W_N(F_m, G_n)| \leq K$$

for G_n . When $\Psi(t) = \sqrt{t(1-t)}$, this inequality becomes

$$[G_n(x) - F_m(x)]^2 \leq \frac{K^2}{M} [\lambda F_m(x) + (1-\lambda)G_n(x)][1 - \{\lambda F_m(x) + (1-\lambda)G_n(x)\}]$$

for x such that $a \leq F_m(x) \leq b$.

Let $c = K^2/M$, $u = F_m(x)$ and $v = G_n(x)$, then the inequality can be written as

$$d(v) \overset{\text{def}}{=} (1+c(1-\lambda)^2)v^2 - [2u-c(1-\lambda)(2\lambda u-1)]v+u^2-c\lambda u+c\lambda^2u^2 \leq 0 .$$

Since the coefficient of v^2 is positive, $d(v) \leq 0$ if and only if v is between the two real roots of the equation $d(v) = 0$. These roots are

$$h^{\pm}(u) \stackrel{\text{def}}{=} \frac{u+\frac{1}{2}c(1-\lambda)(1-2\lambda u) \pm\frac{1}{2}\sqrt{c^2(1-\lambda)^2+4cu(1-u)}}{1+c(1-\lambda)^2} \tag{7}$$

It follows that with probability $(1-\alpha)$, $G_n(x)$ is in the band

$$h^-(F_m(x)) \leq G_n(x) \leq h^+(F_m(x))$$

for all $x \in \{x : a \leq F_m(x) \leq b\}$.

Applying Proposition 1, we have shown

Remark 2:

Let $P_{F=G}(W_N \leq K) = 1 - \alpha$, then the level $(1-\alpha)$, simultaneous, confidence band for $\Delta(x)$ based on W_N with $\psi(t) = \sqrt{t(1-t)}$ is

$$[G_n^{-1}(h^-(F_m(x))) - x \quad , \quad G_n^{-1}(h^+(F_m(x))) - x),$$

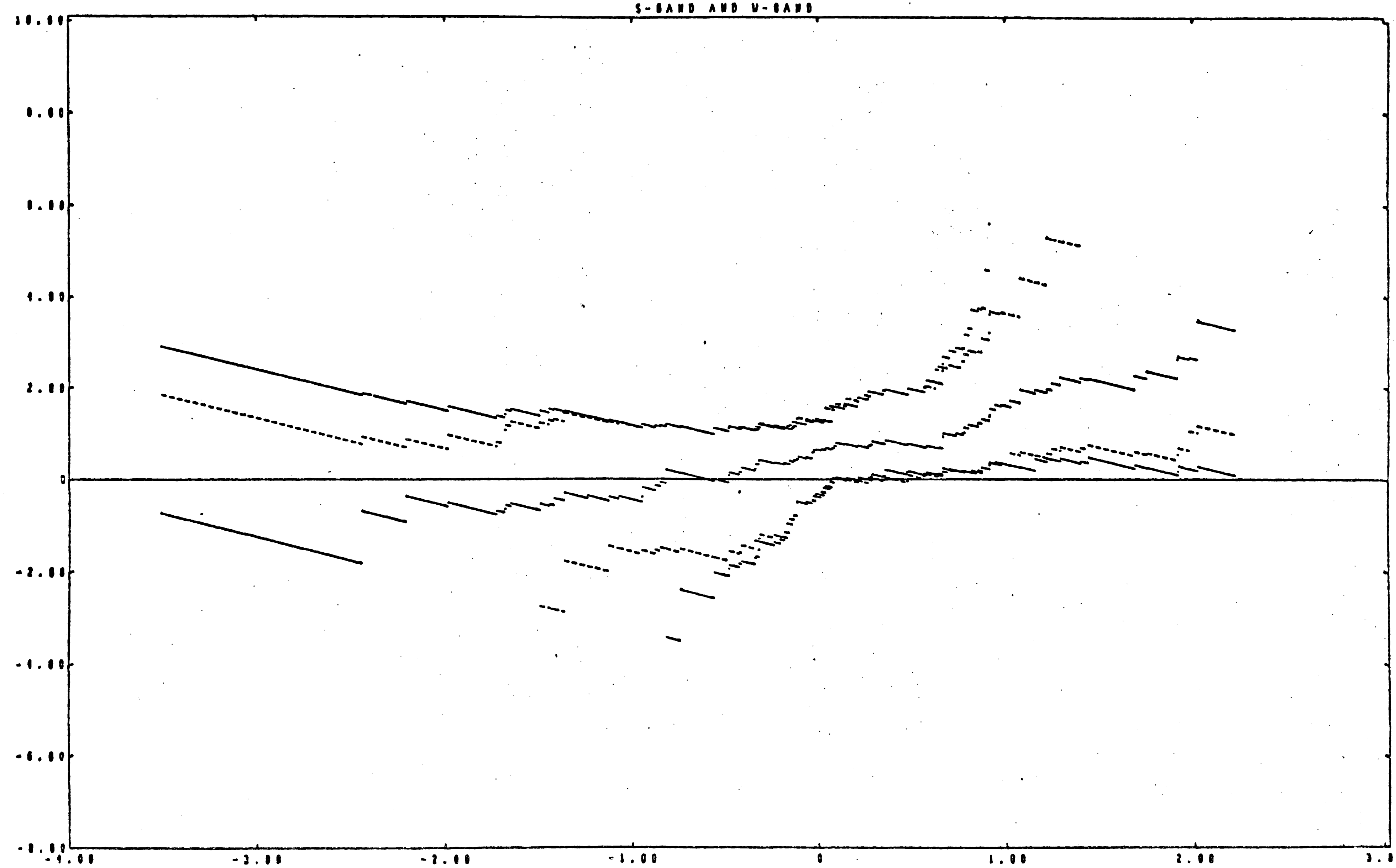
$x \in \{x : a \leq F_m(x) \leq b\}$.

We refer to this band as the W band and write it $[\underline{\Delta}_W(x) , \overline{\Delta}_W(x))$. As with the S band, it is computed from the order statistics by using (5). Monte Carlo values of K are given by Canner (1975) when $a = 1 - b = 0$. Fig. 1 gives this band (dotted lines) for X and Y samples from N(0,1) and N(1,4) distributions, respectively. Here $m = n = 100$, $\alpha = .05$ and $K = 3.02$ is obtained from Canner.

For $a > 0$, $b < 1$, asymptotic critical values K can be obtained from Borokov and Sycheva (1968).

Figure 1.

The estimate $\hat{\Delta}(x)$ and the level .95 S (solid lines) and W (dotted lines) bands.



A third band can be obtained by considering the Renyi statistic

$$R_N = \sqrt{M} \sup_{x \in \{x : F_m(x) \geq c\}} \frac{|F_m(x) - G_n(x)|}{H_N(x)} .$$

This statistic is a possibility when one wants to give more weight to smaller x's . If the X's and Y's are lifetimes, small x's correspond to high risk members of the population. If one wants a band which is accurate for these x's , the band based on R_N could be considered. The inversion of this statistic is straight forward.

Let r denote the level α critical value for R_N and define

$$h_R^{\pm}(u) = \left(\frac{1 \pm \lambda r/\sqrt{M}}{1 \mp (1-\lambda)r/\sqrt{M}}\right)u$$

then the <u>R band</u> $[\underline{\Delta}_R(x), \overline{\Delta}_R(x))$ is obtained by substituting $h_R^{\pm}$ for $\underline{h}$ and $\overline{h}$ in (5) . x is required to be in $\{x : F_m(x) \geq c\}$.

Asymptotic critical values r can be obtained from the tables of Renyi (1953).

3. COMPARISON OF THE NONPARAMETRIC BANDS.

We compare the bands in terms of their widths and their limiting widths

$$w_{M,\alpha}(x) = \sqrt{M}[G_n^{-1}(\overline{h}(F_m(x))) - G_n^{-1}(\underline{h}(F_m(x)))]$$

and

$$w_\alpha(x) = \lim_{M\to\infty} w_{M,\alpha}(x) .$$

When computing this limit, it is convenient to introduce the notation $u = F_m(x)$ and $\epsilon = M^{-\frac{1}{2}}$. Moreover, to emphasize the dependence on ϵ, we write $\underline{h}(\epsilon,u)$ for $\underline{h}(u)$ and $\overline{h}(\epsilon,u)$ for $\overline{h}(u)$. Now

$$\begin{aligned} w_{M,\alpha} = {} & \epsilon^{-1}[G^{-1}(\overline{h}(\epsilon,u)) - G^{-1}(u)] - \epsilon^{-1}[G^{-1}(\underline{h}(\epsilon,u)) - G^{-1}(u)] \\ & + \epsilon^{-1}[G_n^{-1}(\overline{h}(\epsilon,u)) - G^{-1}(\overline{h}(\epsilon,u))] \\ & - \epsilon^{-1}[G_n^{-1}(\underline{h}(\epsilon,u)) - G^{-1}(\underline{h}(\epsilon,u))] \\ & + \epsilon^{-1}[G^{-1}(u) - G^{-1}(u)] \\ & \overset{\text{def}}{=} \sum_{i=1}^{5} I_i . \end{aligned}$$

The limit of $w_{M,\alpha}(u)$ is infinite unless G is strictly increasing at $G^{-1}(u)$. We make this assumption, thus $I_5 = 0$.

Suppose furthermore that G has a continuous, nonzero derivative g. Using a result of Doksum (1974) and a random change of time argument (e.g. Billingsley (1968), pp.144-6), we find that I_3+I_4 converges in law to $U(F(x))/g(\Delta(x)+x) - U(F(x))/g(\Delta(x) + x)$ on every interval $[F^{-1}(\delta), F^{-1}(1-\delta)]$, $0 < \delta < \frac{1}{2}$, where U denotes the Brownian Bridge on $D[0,1]$.

It follows that $I_3 + I_4$ converges in probability to zero. If we add the assumption that $\underline{h}(\epsilon,u)$ and $\overline{h}(\epsilon,u)$ converge to u as $\epsilon \to 0^+$ and that $\underline{h}(\epsilon,u)$ and $\overline{h}(\epsilon,u)$ have continuous (in u) right hand derivatives $\underline{h}'(0,u)$ and $\overline{h}'(0,u)$ with respect to ϵ at $\epsilon = 0$, then we can compute the limits of I_1 and I_2 as follows

$$\lim_{\epsilon\to 0^+} I_1 = \lim_{\epsilon\to 0^+} \frac{G^{-1}(\overline{h}(\epsilon,u)) - G^{-1}(\overline{h}(0,u))}{\epsilon} = \frac{\overline{h}'(0,F(x))}{g(G^{-1}(F(x)))}$$

$$\lim_{\epsilon\to 0^+} I_2 = \frac{\underline{h}'(0,F(x))}{g(G^{-1}(F(x)))}$$

To summarize, we have

<u>Proposition 2:</u>

Under the above stated conditions, the asymptotic widths of the bands (4) are given by

$$w_\alpha(x) = \frac{\overline{h}'(0,F(x)) - \underline{h}'(0,F(x))}{g(G^{-1}(F(x))}\ .$$

Let the <u>asymptotic relative efficiency</u> of two bands be the square of the ratio of the reciprocals of their asymptotic widths. Thus for two bands based on functions $\underline{h}_1,\overline{h}_1$ and $\underline{h}_2,\overline{h}_2$, we have that the efficiency is

$$e_{h_1,h_2}(x) = \left[\frac{\overline{h}_2'(0,F(x)) - \underline{h}_2'(0,F(x))}{\overline{h}_1'(0,F(x)) - \underline{h}_1'(0,F(x))}\right]^2 .$$

If we think of this efficiency as a function of the quantiles $x_q = F^{-1}(q)$ of F, we see that it is independent of the form of F and G.

In the case of the S band, $\underline{h}(\epsilon,u) = u-K_{S,\alpha}\epsilon$ and $\bar{h}(\epsilon,u) = u+K_{S,\alpha}\epsilon$ thus its asymptotic width is

$$w_{S,\alpha}(x_q) = \frac{2K_{S,\alpha}}{g(G^{-1}(q))} , \quad 0 < q < 1 .$$

In the case of the W band, we replace c by $K^2\epsilon^2$ in (7) and differentiate with respect to ϵ to obtain $\underline{h}'(0,u)$ and $\bar{h}'(0,u)$. This yields

$$w_{W,\alpha}(x_q) = \frac{2K\sqrt{q(1-q)}}{g(G^{-1}(q))} , \quad a \leq q \leq b$$

as the asymptotic width of the W band. Thus the asymptotic efficiency of the S band to the W band is

$$e_{S,W}(x_q) = \left(\frac{K_{W,\alpha}}{K_{S,\alpha}} \right)^2 q(1-q) , \quad a \leq q \leq b$$

where $K_{S,\alpha}$ and $K_{W,\alpha}$ denote the asymptotic critical values of the D_N and W_N statistics, respectively.

Using Borokov and Sycheva, we obtain the following very close but slightly conservative approximation to the asymptotic $K_{W,\alpha}$.

Table 1

ASYMPTOTIC CRITICAL VALUES OF W_N. b =1 - a

a \ α	.2	.1	.04	.02
.25	2.109	2.482	2.879	3.138
.1	2.471	2.758	3.091	3.318

Combining this with the asymptotic tables for D_N, we obtain the following tables of efficiencies.

Table 2

THE ASYMPTOTIC EFFICIENCY $e_{S,W}(x_q)$ of the S BAND TO THE W BAND WHEN α = .1

a = 1 - b = .25

α \ q	.25	.30	.35	.40	.45	.50
.1	.78	.87	.94	1.00	1.02	1.04

$a = 1 - b = .1$

α \ q	.1	.2	.3	.4	.5
.1	.46	.82	1.07	1.23	1.28

The values for $q > .5$ are the same as those for $1-q$. Different values of α yielded similar results. We find that the S band is better in the center of the X distribution at the expense of being worse in the tails up to $\pm x_a$. Technically, the asymptotic efficiency is ∞ for x outside $\pm x_a$. However, the S band is only valid for $F_m(x)$ in $(K_{S,\alpha}/\sqrt{M}, 1-K_{S,\alpha}/\sqrt{M})$. When m=n and α=.01, $K_{S,\alpha}/\sqrt{M} \geq .1$ for $n \leq 531$, while for $\alpha = .1$, $K_{S,\alpha}/\sqrt{M} \geq .1$ for $n \leq 297$.

Canner (1975) gives Monte Carlo critical values of W_N with $a = 1 - b = 0$ when n=m=2000(α=.05) and n=m=1000(α=.01). Using these in place of asymptotic critical values, we obtain

Table 3

APPROXIMATION TO THE ASYMPTOTIC EFFICIENCY

$a = 1 - b \cong 0$

α \ q	.01	.05	.1	.2	.3	.4	.5
.05	.06	.26	.50	.89	1.17	1.34	1.39
.01	.05	.24	.46	.82	1.08	1.24	1.29

These values are close to the values of table 2 with $a = .1$.

We also computed the efficiencies for finite sample sizes (reciprocal ratios of widths for actual samples) using Canner's critical values and computer generated samples. Some results are given below for X samples from a N(0,1) distribution and Y samples from a $N(1,\sigma_2^2)$ distribution.

Table 4.

Finite sample size efficiency of the S band to the W band. $\alpha = .05$.

$m = n = 50$.

σ_2 \ q	.1	.2	.3	.4	.5	.6	.7	.8
.5	∞/∞	0	0	.61	1.01	1.00	.76	∞/∞
2.0	∞/∞	0	0	.61	1.18	1.00	.66	∞/∞

$m = n = 100$

σ^2 \ q	.05	.1	.2	.3	.4	.5	.6	.7	.8	.9
.5	∞/∞	0	0	.87	.97	1.18	1.18	.80	.65	∞/∞
2.0	∞/∞	0	0	.80	.90	1.14	1.19	1.09	.56	∞/∞

These results are qualitative close to the asymptotic results bur favor the W band more. They vary little from sample to sample or from distribution to distribution.

Taken together, the tables show that in terms of width the W band is preferable to the S band. The S band is better in the center of the X distribution at the expense of being much worse in the tails. This is also clear from Fig. 1. It is interesting to note that in this figure, the W band leads correctly to the rejection of a linear model while the S band does not (see (iii), section 1). The S band has the advantage of being simpler and its critical values are more extensively tabulated. It is preferable if the central part of the X distribution is of more interest than the tails.

For the R-band, Proposition 2 yields the asymptotic width

$$w_{R,\alpha}(x_q) = \frac{2K_{R,\alpha}q}{g(G^{-1}(q))}, \quad c < q < 1$$

where $K_{R,\alpha}$ is the asymptotic critical value of the Renyi statistic R_N. From Renyi (1953), we obtain

Table 5

ASYMPTOTIC CRITICAL VALUES OF R_N

c	.2	.1	.05
$K_{R,\alpha}$	4	6	8.5
$1-\alpha$	.9082	.9081	.8977

Table 6

ASYMPTOTIC EFFICIENCY OF THE R BAND TO THE W BAND

$a = 1 - b = .1$, $c = .05$, $\alpha = .1$

q	.05	.1	.2	.3	.4	.5	.6
e	∞	1.05	.47	.28	.18	.11	.08

$a = 1 - b = .1$, $c = .1$

q	.1	.2	.3	.4	.5	.6	.7
e	2.08	.96	.45	.35	.22	.15	.10

$a = 1 - b = .25$, $c = .2$

q	.2	.25	.3	.4	.5	.6	.7
e	∞	1.15	.89	.58	.39	.26	.17

It is apparent that the W-band is preferable to the R band except when only small x_q are of interest.

4. THE LOCATION-SCALE MODEL.

In this section, we consider the model where

$$F(x) = H(\frac{x-\mu_1}{\sigma_1}) \quad , \quad G(y) = H(\frac{y-\mu_2}{\sigma_2})$$

for some continuous distribution function H .
Then $G^{-1}(u) = \mu_2 + \sigma_2 H^{-1}(u)$ and

$$\Delta(x) = \mu_2 + \frac{\sigma_2}{\sigma_1}(x-\mu_1) - x .$$

In this model it is common to treat $\mu_2-\mu_1$ as the parameter of interest. However, as seen in section 1 , $\Delta(x)$ will yield additional information about how the populations differ.

Since $\Delta(x)$ is linear in x , a simultaneous band can be constructed by specifying intervals of values for $\Delta(x)$ at just two points x .
Thus if $x_1 < x_2$ and

$$P(\underline{T} \leq \Delta(x_1) \leq \bar{T} \quad , \quad \underline{S} \leq \Delta(x_2) \leq \bar{S}) = 1 - \alpha$$

then for $x \in [x_1,x_2]$, the upper boundary would consist of the line connecting the points $(x_1,\bar{T})$ and $(x_2,\bar{S})$, while the lower boundary would be the line connecting $(x_1,\underline{T})$ and $(x_2,\underline{S})$. For $x > x_2$, the upper boundary would be the line through the points $(x_1,\underline{T})$ and $(x_2,\bar{S})$, and so on. x_1 and x_2 may be random, in fact in the following, they are order statistics.

Proposition 3.

Suppose that a location-scale model holds. Let r_k , i_k , s_k , $k = 1,2$, be integers such that

$$P_{F=G}(Y(r_k) < X(i_k) < Y(s_k)\ ,\ k=1,2) = 1 - \alpha \qquad (8)$$

Then a simultaneous $(1-\alpha)$ 100% confidence band for $\Delta(x)$ is determined by

$$Y(r_k) - X(i_k) < \Delta(X(i_k)) < Y(s_k) - X(i_k), k=1,2 \quad (9)$$

Proof:

For arbitrary continuous F and G , $\Delta(X) + X$ and Y have the same distribution. Thus

$$P(Y(r_k) < \Delta(X(i_k)) + X(i_k) < Y(s_k)\ ,\ k=1,2) = 1 - \alpha$$

and the result follows.

The choice of integers r_k , i_k , s_k satisfying (8) could be made using the bivariate hypergeometric distribution, although convenient tables do not seem to be available.

Alternatively, Switzer (1975) has suggested a conservative procedure using the Bonferroni inequality and the hypergeometric distribution. A third possibility, when sample sizes are large, is to use a bivariate normal approximation. To do this, let Z_1 = No. of $Y_j \leq X(i_1)$ and Z_2 = No. of $Y_j > X(i_2)$.
Then the probability in (8) equals
$P_{F=G}\ [r_1 \leq Z_1, \leq s_1 - 1\ ,\ n - s_2 + 1 \leq Z_2 \leq n - r_2]$ and can be approximated using

Lemma 1.

For $0 < \beta_1 < \beta_2 < 1$, let $i_k = [m\beta_k] + 1$, $k=1,2$, let $\xi_1 = \beta_1$, $\xi_2 = 1 - \beta_2$, and let

$$V_k = \sqrt{M}(\frac{Z_k}{m} - \xi_k)/\sqrt{\beta_k(1-\beta_k)} .$$

If $F = G$, then (V_1,V_2) has a standard bivariate normal limiting distribution with correlation $\rho = -[\beta_1(1-\beta_2)/(1-\beta_1)\beta_2]^{\frac{1}{2}}$.

The lemma is proved by noting that the conditional distribution of (Z_1,Z_2) given $X(i_1)$, $X(i_2)$ is trinomial and after standardization, asymptotically bivariate normal. Since $E(Z_k|X(i_k))$, $i=1,2$, is asymptotically normal, an application of Sethurarman's theorem gives the result.

The following remark gives a particular application of the lemma.

Remark 3.

Let $i_1 = [m\beta] + 1$ and $i_2 = [m(1-\beta)] + 1$ for some $\beta \in (0,\frac{1}{2})$. Then an asymptotic $(1-\alpha)$ 100% simultaneous confidence band for $\Delta(x)$ is determined by (9) with

$$\frac{r_1-\frac{1}{2}}{n} = \beta - c_\alpha\sqrt{\beta(1-\beta)}/\sqrt{M} ,$$

$$\frac{s_1-\frac{1}{2}}{n} = \beta + c_\alpha\sqrt{\beta(1-\beta)}/\sqrt{M} , \qquad (10)$$

$$r_2 = n+1-s_1 \quad \text{and} \quad s_2 = n+1-r ,$$

where c_2 satisfies $P[|V_k| \leq c_2 \ ; \ k=1,2] = 1 - \alpha$ for (V_1,V_2) standard bivariate normal with correlation $\rho = - \beta/(1-\beta)$.

If $L_M(x)$ denotes the width of this band at x , and if H has a density h , then

$$\sqrt{M}\, L_M(x) \xrightarrow{P} \frac{2c_\alpha \sqrt{\beta(1-\beta)}\,\sigma_2}{(z_{1-\beta}-z_\beta)} \left\{ \frac{|z_p - z_\beta|}{h(z_{1-\beta})} + \frac{|z_p - z_{1-\beta}|}{h(z_\beta)} \right\}$$

$$= L(x) \text{ , say}$$

for $x = \sigma_1 z_p + \mu_1$ with $\beta < p < 1-\beta$ and z_p the p^{th} quantile of H .

If the density $h(x)$ is symmetric about 0 , then

$$L(x) = 2c_\alpha \sqrt{\beta(1-\beta)}\,\sigma_2 / h(z_\beta) \quad , \quad \beta < p < 1-\beta \ .$$

We call the band determined by (9) and (10) the O band (order statistics band) and compare it with the W band of section 2 . In the location scale model, the asymptotic width of the W band is $2K_{W,\alpha}\sqrt{p(1-p)}\,\sigma_2/h(z_p)$. When H is normal, we obtain

Table 7.

The asymptotic efficiency of the O band to the W band with $a = 1-b = \beta$ in the normal model . $\alpha = .1$.

β \ p	.10	.20	.25	.30	.40	.50
.10	1.99	1.39	1.27	1.19	1.10	1.08
.25			1.63	1.53	1.41	1.38

These efficiencies will not be much different for other reasonable "bell-shaped " h . Thus we conclude that if a location-scale model can be assumed, a considerable gain in efficiency is possible by using the O-band provided only that $h(z_\beta)$ and $h(z_{1-\beta})$ are not to close to zero.

5. THE NORMAL MODEL.

If we let H in the location-scale model be $N(0,1)$, then we have the Behrens-Fisher model. In this case we can write

$$\Delta(x) = ax + b - x$$

where

$$a = \frac{\sigma_2}{\sigma_1} \quad , \quad b = \mu_2 - \frac{\sigma_2}{\sigma_1}\mu .$$

We will construct the likelihood ratio test for the hypothesis $H_0 : a=a_0 , b=b_0$ for fixed $a_0 \epsilon (0,\infty)$ and $b_0 \epsilon (-\infty, \infty)$. The collection of all (a_0,b_0) that is accepted by this test for a given set of data provides a confidence region for (a,b) that is an ellipsoid. This ellipsoid will be translated into a likelihood ratio simultaneous confidence band for $\Delta(x)$.

If L denotes the likelihood function, then

$$L \propto \sigma_1^{-m}\sigma_2^{-n} \exp \{- \frac{1}{2\sigma_1^2} \sum_1^m (x_i-\mu_1)^2 - \frac{1}{2\sigma_2^2} \sum_1^n (y_j-\mu_2)^2\}$$

where $\propto$ denotes proportional to.

The maximum for unrestricted $\mu_1, \mu_2, \sigma_1, \sigma_2$ is well known. Substituting $\sigma_2^2 = a_o^2 \sigma_1^2$ and $\mu_2 = b_o + a_o \mu_1$, the maximum of L under H_o occurs at

$$\hat{\mu}_1 = \frac{\hat{a}_o m\bar{x} + n(\bar{y} - b_o)}{a_o N}$$

$$\hat{\sigma}_1^2 = \frac{1}{N}\{\Sigma(x_i - \hat{\mu}_1)^2 + \frac{1}{a_o^2}\Sigma(y_j - b_o - a_o\hat{\mu}_1)^2\}$$

$$= \frac{1}{N}\{ms_1^2 + \frac{ns_2^2}{a_o^2} + M\frac{(\bar{y} - b_o - a_o\bar{x})^2}{a_o^2}\}$$

where $(\bar{x}, \bar{y}, S_1^2, S_2^2)$ denotes the usual unrestricted maximum likelihood estimate of $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$. It follows that the likelihood ratio is

$$\Lambda_N = \frac{\sup_{H_o} L}{\sup L} = \frac{(S_1^2)^{\frac{1}{2}m}(S_2^2)^{\frac{1}{2}n}}{(\hat{\sigma}_1^2)^{\frac{1}{2}m}(\sigma_o^2\hat{\sigma}_1^2)^{\frac{1}{2}n}}$$

$$= \frac{(a_o^2 S_1^2)^{\frac{1}{2}m}(S_2^2)^{\frac{1}{2}n}}{[\lambda a_o^2 S_1^2 + (1-\lambda)S_2^2 + \lambda(1-\lambda)(\bar{y} - b_o - a_o\bar{x})^2]^{\frac{1}{2}N}}$$

where $\lambda = \frac{m}{N}$.

The space of $(\mu_1, \mu_2, \sigma_1, \sigma_2)$ with $\sigma_2 = a_o \sigma_1$ and $\mu_2 = b_o + a_o \mu_1$ is linear with 2 dimensions. Thus, under H_o, by classical results, $-2 \log \Lambda_N$ has a limiting $\chi_{4-2}^2 = \chi_2^2$ distribution.

Let $x^2(1-\alpha)$ denote the $(1-\alpha)$th quantile of the χ_2^2 distribution and let $K_\alpha = \exp\{x^2(1-\alpha)/N\}$ then the acceptance region of the test is

$$\frac{\lambda a_o^2 S_1^2 + (1-\lambda)S_2^2 + \lambda(1-\lambda)(\bar{y}-b_o-a_o\bar{x})^2}{(a_o^2 S_1^2)^{\lambda}(S_2^2)^{(1-\lambda)}} \leq K_\alpha .$$

When $n = m, \lambda = \frac{1}{2}$, this simplifies to

$$T \overset{\text{def}}{=} \frac{\frac{1}{2}a_o^2 S_1^2 + \frac{1}{2}S_2^2 + \frac{1}{4}(\bar{y}-b_o-a_o\bar{x})^2}{a_o S_1 S_2} \leq K_\alpha .$$

Write $b_o' = b_o + a_o\bar{x}$, then the collection of (a_o, b_o') for which H_o is accepted can be written as the inside of an ellipse as follows:

$$2(S_x a_o - K_\alpha S_y)^2 + (\bar{y}-b_o')^2 \leq 2S_y^2(K_\alpha^2-1) \qquad (11)$$

Let E denote this ellipse and let

$$\delta^-(x) = \inf\{a_o x + b_o : (a_o, b_o') \in E\} - x$$

$$\delta^+(x) = \sup\{a_o x + b_o : (a_o, b_o') \in E\} - x$$

then $[\delta^-(x), \delta^+(x)]$ is the level $(1-\alpha)$ L.R. confidence band for $\Delta(x)$. $\delta^{\pm}$ is given in

<u>**Proposition 4.**</u> When $n=m$, the level $(1-\alpha)$ likelihood ratio confidence band for $\Delta(x)$ in normal models is given by

$$\delta^{\pm}(x) = \bar{Y} + K_\alpha S_2\left(\frac{x-\bar{X}}{S_1}\right) - x \pm S_2\sqrt{K_\alpha^2-1}\sqrt{2+\frac{(x-\bar{X})^2}{S_1^2}} .$$

Proof :

We will use the Cauchy-Schwartz inequality in the form

$$|\Sigma\, \alpha_i w_i| \leq \sqrt{\Sigma\, \alpha_i^2\, \Sigma w_i^2} \qquad (12)$$

with equality iff w_i is proportional to α_i.

Simply note that (11) is equivalent to

$w_1^2 + w_2^2 \leq d^2$ where

$w_1 = \sqrt{2}(S_1 a - K_\alpha S_2)$

$w_2 = \bar{y}-b-a\bar{x}$ and $d^2 = 2S_2^2(K_\alpha^2-1)$.

If we apply (12) with $\alpha_1 = -(x-\bar{x})/\sqrt{2}S_1$ and $\alpha_2 = 1$, the result follows.

Remark 4 : The above proof is very similar to the derivation of Scheffé's simultanous confidence intervals for contrasts. The interpretation is also similar: If the likelihood ratio test of H_0 : "$\Delta(x) \equiv 0$ for all x"(i.e. a=1,b=0) rejects, then for some x the band does not contain 0 and vice versa.

When $n \neq m$, we can use the maximum likelihood estimate

$$\hat{\Delta}(x) = \bar{y} + \frac{S_2}{S_1}(x-\bar{x}) - x$$

of $\Delta(x)$ to obtain a confidence band for $\Delta(x)$. Let

$$T_N(x) = \sqrt{M}[\hat{\Delta}(x) - \Delta(x)] .$$

By using a Taylor expansion of $T_N(x)$ in terms of $\bar{X}, \bar{Y}, S_1$ and S_2 , we find that $T_N(x)$ converges in law to

$$T(x) \overset{def}{=} \sigma_2[V_1 + tV_2/\sqrt{2}]$$

where $t = (x-\mu_1)/\sigma_1$ and V_1, V_2 are independent standard normal variables.

Write $\tau^2 = \text{Var}\ (T(x)) = \sigma_2^2[1+\frac{1}{2}t^2]$ and

$$\hat{\tau}^2 = S_2^2\ (1+\tfrac{1}{2}(x-\bar{x})/S_1)$$

then we can use $\sup_x |T_N(x)|/\hat{\tau}$ as a pivot and obtain the following <u>M.L. band</u>

<u>Proposition 5</u>.

An asymptotic $100(1-\alpha)\%$ simultaneous confidence band for $\Delta(x)$ is given by

$$\Delta(x) \in \bar{Y} + \frac{S_2}{S_1}(x-\bar{x}) - x \pm x(1-\alpha)[\{1+\tfrac{1}{2}(x-\bar{x})^2/S_1^2\}/M]^{\frac{1}{2}}$$

for all x , where $x^2(1-\alpha)$ is the $(1-\alpha)$th quantile of the χ_2^2 distribution.

<u>Proof :</u> $T_N(x)/\hat{\tau}$ converges in law to the process $(V_1 + tV_2/\sqrt{2}) / \sqrt{1 + \frac{1}{2}t^2}$. Hence $\sup_x |T_N(x)/\hat{\tau}|$ converges in law to

$$\sup_t |(V_1+tV_2/\sqrt{2}) / \sqrt{1 + \tfrac{1}{2}t^2}| = \sup_t |l(t)| \quad \text{(say)}$$

By considering the equation $l'(t) = 0$, we find that for almost all (V_1, V_2) , the maximum of $|l(t)|$ is attained at $t = \sqrt{2}V_2/V_1$, a Cauchy random variable. Thus the maximum is $\sqrt{V_1^2 + V_2^2}$, which is the square root of a χ_2^2 variable. The result follows.

By standard asymptotic theory, the L.R. and M.L. bands should be asymptotically equivalent. This can be established directly when $n=m$ by using the first 2 terms in the Taylor expansion of e^z about $z=0$ in the expression $k_\alpha = \exp\{x^2(1-\alpha)/N\}$. The L.R. band is preferable since it is based on a more accurate approximation. The above expansion also yields :

<u>Proposition 6.</u> The L.R. and M.L. bands both have asymptotic width

$$2\sigma_2\, x(1-\alpha)\sqrt{1 + \tfrac{1}{2}t^2}$$

where $t = (x-\mu_1)/\sigma_1$.

It is interesting to compare this asymptotic width with the asymptotic widths of the general methods of the previous sections to find out how much these general methods loose if in fact the correct model is normal.

We see that the asymptotic relative efficiency of the S band to the L.R. band in normal models is

$$e_{S,L}(x_q) = \lim_{M\to\infty} \frac{[\text{Width L band}]^2}{[\text{Width S band}]^2}$$

$$= \frac{\varphi^2(t)\, x^2(1-\alpha)\, (1+\frac{1}{2}t^2)}{K_{S,\alpha}^2}$$

$$= \frac{(1+\frac{1}{2}t^2)\, x_2^2(1-\alpha)}{2\pi\, e^{t^2} K_{S,\alpha}^2}$$

where $\varphi(t)$ denotes the standard normal density. Similar expressions hold for the W and O bands. Some numerical results are

Table 8.

Asymptotic efficiencies with respect to the L.R. band in normal models. $t = \Phi^{-1}(p)$, $\alpha =$ .10

A. The W-band with $a = 1-b$.

a \ p	.10	.20	.25	.30	.40	.50
.10	.38	.40	.40	.40	.39	.39
.25			.49	.49	.48	.48

B. The S-band.

p	.10	.20	.30	.40	.50
	.17	.33	.43	.48	.49

C. The O-band.

β \ p	.10	.20	.25	.30	.40	.50
.10	.75	.56	.51	.48	.42	.42
.25			.80	.75	.68	.66

The efficiency of the S band is surprisingly low, much smaller than the familiar $\pi/2 = .64$.

We also give some finite sample size "efficiencies" computed for the same $N(0,1)$ and $N(1,\sigma_2^2)$ samples as in section 3 .

Table 9.

Finite sample size efficiency of the W band with $a = 1-b \cong 0$ to the L.R. band. $\alpha = .05$ and $m=n$.

n	σ_2 \ p	.1	.2	.3	.4	.5	.6	.7	.8	.9
50	0.5	0	.31	.34	.46	.63	.62	.44	0	0
	2.0	0	.35	.27	.48	.53	.57	.66	0	0
100	0.5	.31	.33	.28	.46	.73	.51	.50	.31	0
	2.0	.33	.36	.27	.33	.65	.52	.57	.34	0

By multiplying the entries of this table with the entries of table 4, we get the corresponding table for the S band. The asymptotic efficiencies evidently give a good indication of the finite sample size performance of the bands.

The results show that the general bands are quite inefficient if the correct model is normal. On the other hand, the bands designed for the normal model are quite sensitive to the normality assumption in the sense that skewness or high kurtosis in the F and G distributions will alter the level of the band. Also note that the O and L.R. bands can not be used to test whether a location-scale model holds. Finally, the general methods W , S and O have the advantage that they can be applied to censored data.

In figure 2 we give a plot of the L.R. band (solid lines) together with the maximum likelihood estimate (solid line)

$$\hat{\delta}(x) = \bar{Y} + \frac{S_2}{S_1}(x-\bar{x}) - x$$

of $\Delta(x)$ for the same $N(0,1)$ and $N(1,4)$ samples used in figure 1 of section 2 .

The dotted line is the band obtained by inverting the pivot

$$U_N = \sqrt{M}\ \sup_t \left|\Phi\left(\frac{t-\bar{x}}{S_1}\right) - \Phi\left(\frac{\Delta(t)+t-\bar{Y}}{S_2}\right)\right|$$

where Φ is the $N(0,1)$ distribution function.

U_N converges in law to

$$U = \sup_t |\varphi(t)(V_1+tV_2/\sqrt{2})|$$

We have not found the distribution of U and have instead used the bound

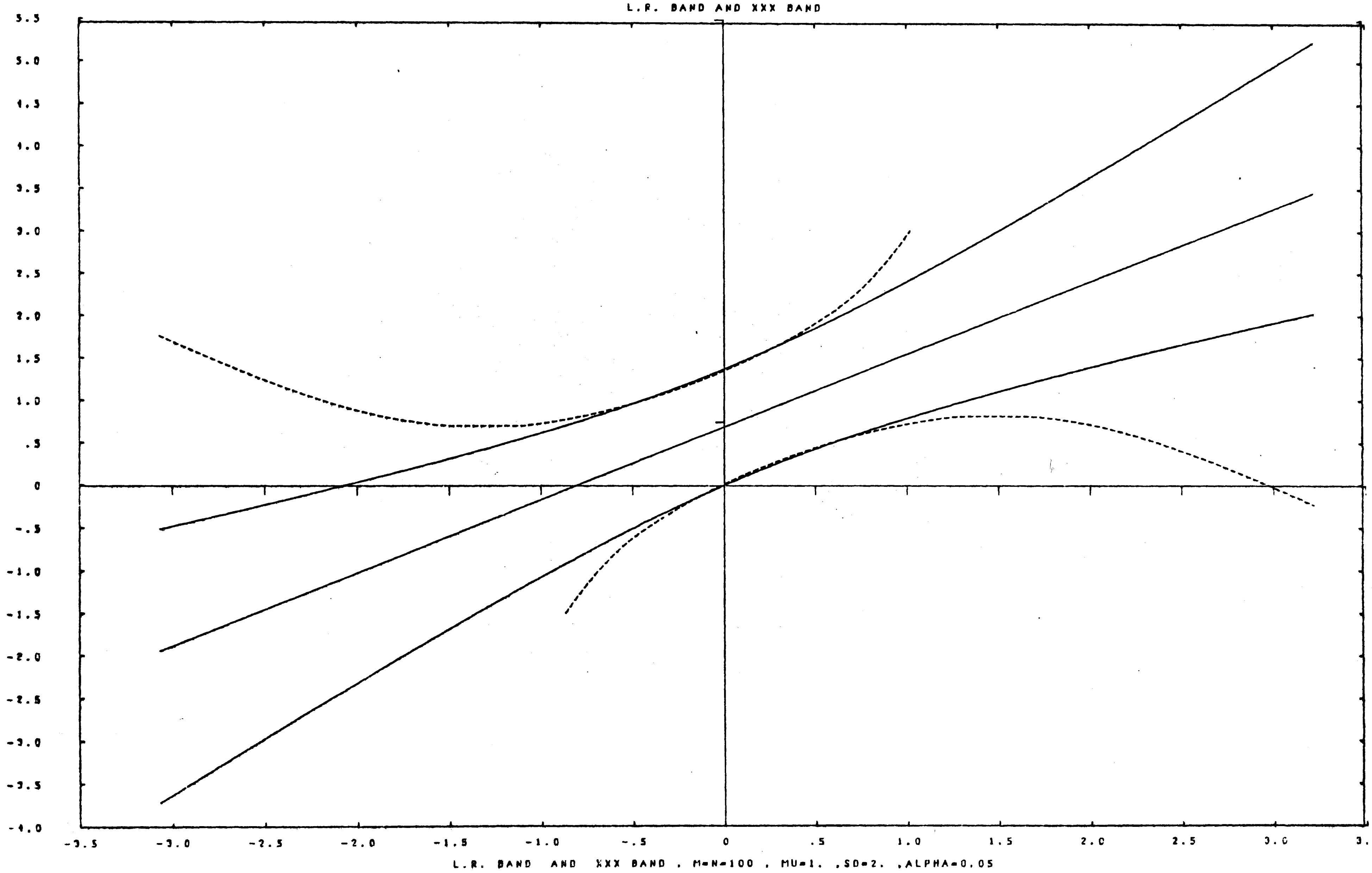
$$U \leq \frac{1}{\sqrt{2\pi}} \max\{|V_1+V_2/\sqrt{2e}|\ ,\ |V_1-V_2/\sqrt{2e}|\}$$

which the plot shows to be barely inadequate.

The resulting band is called the S_2 band.

Figure 2.

The estimate $\hat{\delta}(x)$ and the level .95 L.R.(solid lines) and S_2(dotted lines) bands.



6. AN ILLUSTRATION.

Doksum (1974) gives an example involving experimental data where a linear response function and thus a location-scale model is indicated, and where the response function shows that high risk members of a guinea pig population is affected entirely different by a T.B. dose than low risk members.

Here we give an illustration involving data from an experiment designed to study undesirable effects of ozone,one of the components of California smog.

One group of 22 seventy day old rats were kept in an ozone environment for 7 days and their weight gains y_i noted. Another group of 23 similar rats of the same age were kept in an ozone free environment for 7 days and their weight gains x noted. The results were (furnished courtesy of Brian Tarkington, California Primate Research Center, Univ. of Calif., Davis. From study supported by N.I.H., U.S.P.H.S., ES-0628.)

x	41.0	38.4	24.4	25.9	21.9	18.3	13.1	27.3	28.5	-16.9	26.0	17.4
y	10.1	6.1	20.4	7.3	14.3	15.5	-9.9	6.8	28.2	17.9	-9.0	-12.9

x	21.8	15.4	27.4	19.2	22.4	17.7	26.0	29.4	21.4	26.6	22.7
y	14.0	6.6	12.1	15.7	39.9	-15.9	54.6	-14.7	44.1	-9.0	

x = control and y = treatment

Figure 3 gives the estimate $\hat{\Delta}(x) = G_n^{-1}(F_m(x)) - x$ and the S-band with exact level .90 . The estimate $\hat{\Delta}$ indicates a V-shaped response function, thus moderate weight gains would be reduced most by the ozone treatment. Even though ozone reduces average weight gain, $\hat{\Delta}$ suggests that large weight gains are made even larger!

From the S-band , we can not reject a linear model assumption. This may be because the sample sizes are too small leaving the band too wide. The V-shape of $\hat{\Delta}$ indicates that we can not use any of the narrower location-scale model bands. The upper boundary of the S-band shows that weight gain is reduced significantly for x in the weight gain interval [7.5 , 22.5].

Acknowledgements.

We are indepted to P. Nanopoulus for programing the plots and finite sample size efficiencies. This paper was prepared with the partial support of National Science Foundation Grant GP-33697X2. Doksum's work was partially done while visiting the Department of Mathematics, University of Oslo, while Siever's work was done while visiting the Department og Statistics, University of California, Berkeley.

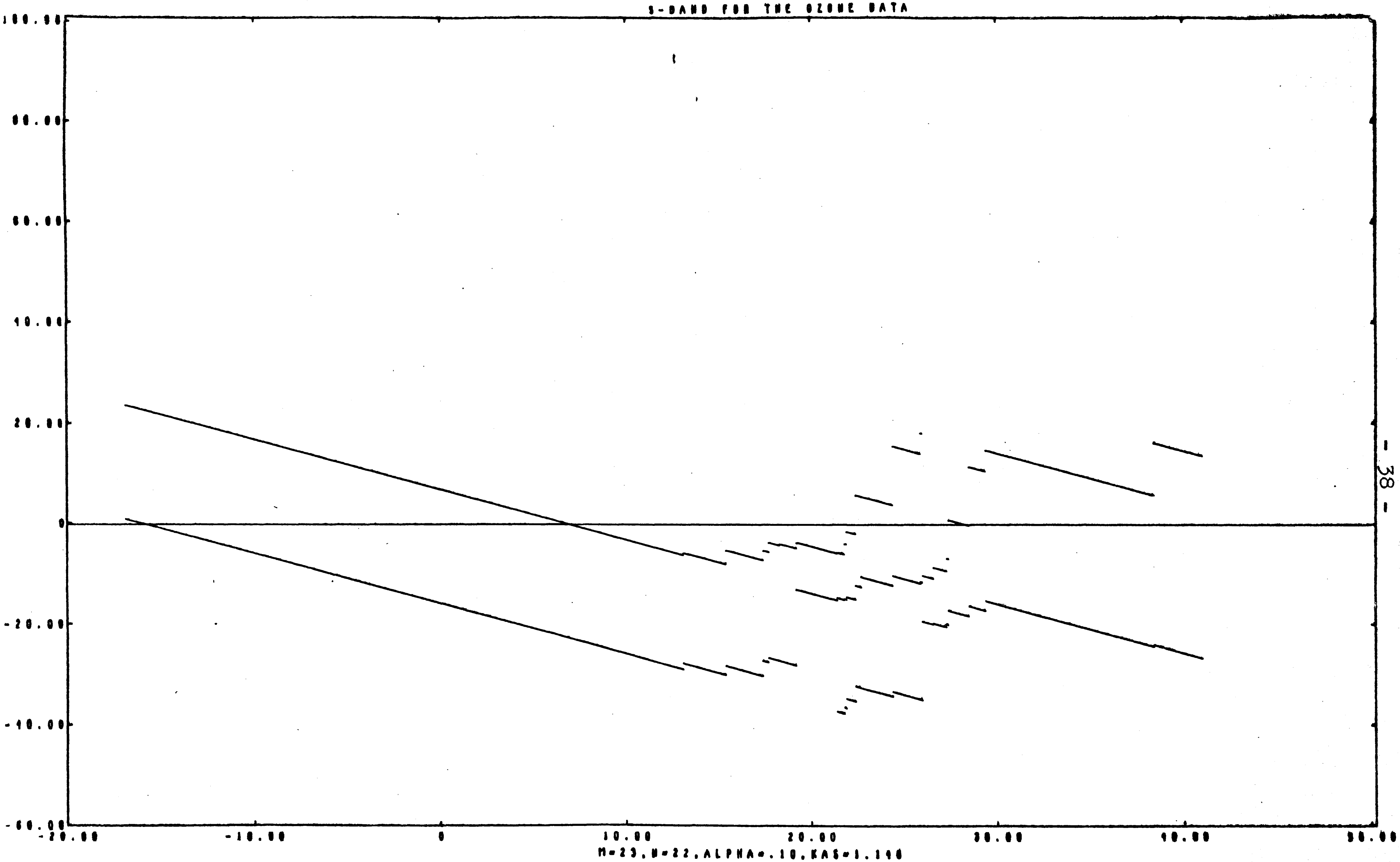


Figure 3.

The estimate $\hat{\Delta}(x)$ and the level .90 S-band for the response function in the ozone experiment.

Fig. 1. The estimate $\hat{\Delta}(x)$ and the level .95 S (solid lines) and W (dotted lines) bands.

Fig. 2. The estimate $\hat{\delta}(x)$ and the level .95 L.R. (solid lines) and S_2 (dotted lines) bands.

Fig. 3. The estimate $\hat{\Delta}(x)$ and the level .90 S-band for the response function in the ozone experiment.

REFERENCES

BILLINGSLEY, P. (1968). Convergence of Probability Measures. New York : Wiley.

BOROKOV, A.A. & Sycheva, N.M. (1968). On asymptotically optimal non-parametric criteria. Th.Prob. and Appl. 13, 359-393.

CANNER, P.L. (1975). A simulation study of one- and two-sample Kolmogorov-Smirnov statistics with a particular weight function. J.Am.Statist.Assoc. 70, 209-211.

DOKSUM, K.A. (1974). Empirical probability plots and statistical inference for nonlinear models in the two sample case. Ann. of Statist. 2, 267-277.

KIM, P.J. & JENNRICH, R.I. (1968). Tables of the upper 2 points of the exact sampling distribution of the two sample Kolmogorov-Smirnov criterion D_{mn} $(m \leq n)$. Biometrika Tables for Statisticians II.

KIM, P.S. & JENNRICH, R.I. (1970). Tables of the exact sampling distribution of the two-sample Kolmogorov-Smirnov criterion $D_{mn}(m \leq n)$. Selected Tables in Math.Statist. 1.

RENYI, A. (1953). On the theory of order statistics. Acta Math. Acad. Sci. Hungar. 4, 191-231.

SIEVERS, G.L. (1975) Confidence bands in non-linear models for the two sample problem. Unpublished paper.

STECK, G.P., ZIMMER, W.J. & WILLIAMS, R.E. (1974), Estimation of parameters in acceleration models. Proc.Annual Reliability and Maintability Symposium, Los Angeles,

SWITZER, P. (1975). Confidence procedures for two sample problems. Biometrika. To appear.

WILK, M.B. & GNANADESIKAN, R. (1968). Probability plotting methods for the analysis of data. Biometrika 55, 1-17.