Plotting with Confidence: Graphical Comparisons of Two Populations

Kjell A. Doksum; Gerald L. Sievers

Biometrika, Vol. 63, No. 3. (Dec., 1976), pp. 421-434.

Stable URL:
http://links.jstor.org/sici?sici=0006-3444%28197612%2963%3A3%3C421%3APWCGCO%3E2.0.CO%3B2-T

Biometrika is currently published by Biometrika Trust.

Biometrika (1976), **63**, 3, *pp.* 421–34
With 2 text-figures
Printed in Great Britain

Plotting with confidence: Graphical comparisons of two populations

BY KJELL A. DOKSUM
Department of Statistics, University of California, Berkeley
AND GERALD L. SIEVERS
Department of Mathematics, Western Michigan University, Kalamazoo

SUMMARY

Statistical methods that give detailed descriptions of how populations differ are considered. These descriptions are in terms 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: Behrens–Fisher model; Confidence bands; Empirical probability plot; Nonlinear model; *Q–Q* plot; Response function; Shift function; 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 response function. Under general conditions it is the only function of x that satisfies $X+\Delta(X) \sim Y$, where $\sim$ denotes distributed as. Thus $\Delta(.)$ 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 shift function.

Assume that F and G are continuous. Let $F^{-1}(u) = \inf\{x : F(x) \geqslant u\}$ be 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 $\Delta(x) = \theta$.

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 & 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 $\{x : \Delta(x) > 0\}$?

This kind of model 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 $\Delta(x) = \theta$, for some θ and all x? (iv) If not, does a shift-scale model hold, i.e. is $\Delta(x) = \alpha + \beta x$, for some α and β and for all x?

These questions can be answered by giving a confidence band $[\Delta_*(x), \Delta^*(x)]$ for $\Delta(x)$, simultaneous for all x. Thus (i) is answered in the affirmative if $\Delta_*(x) > 0$ for all x, (ii) has solution $\{x : \Delta_*(x) > 0\}$, (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 $\Delta_*(x)$.

Such confidence bands have been considered by Doksum (1974), Switzer (1976), and G. L. Sievers in an unpublished report. Switzer and Sievers derive a band based on the two-sample Kolmogorov–Smirnov statistic and thereby obtain the exact confidence coefficient of the band of Doksum (1974). Similar bands have been considered by Steck, Zimmer & Williams (1974) in connection with the 'acceleration function' GF^{-1}. Here we consider bands based on statistics of the form

$$\sup_{a_m \leqslant x \leqslant b_m} |F_m(x) - G_n(x)| / \psi\{H_N(x)\},$$

where $N = m + n$, $\lambda = m/N$ and $H_N(x) = \lambda F_m(x) + (1-\lambda) G_n(x)$.

Efficiency comparisons are made between such bands in terms of squared ratios of widths, and it is found that the choice of $\psi(u) = \{u(1-u)\}^{\frac{1}{2}}$ leads to an efficient band.

If 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 when $n = m$, and it is shown to be much more efficient than the general bands in the normal model. A band based on the maximum likelihood estimate of $\Delta(x)$ is also considered for this 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 the empirical distribution functions F_m and G_n. The key to finding such pivots is 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. Note that

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

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

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

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

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

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 $T(F_m, G_n) \leqslant K$ is equivalent to

$$h_*\{F_m(x)\} \leqslant G_n(x) \leqslant h^*\{F_m(x)\} \tag{2}$$

for all x, for some functions h_* and h^*. Typically these functions are nondecreasing. For instance, let $N = m + n$, $M = mn/N$ and suppose

$$T(F_m, G_n) = D_N = M^{\frac{1}{2}} \sup_x |F_m(x) - G_n(x)|,$$

the Kolmogorov–Smirnov statistic. Then $h_*(x) = x - K/M^{\frac{1}{2}}$ and $h^*(x) = x + K/M^{\frac{1}{2}}$.

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

$$\mathrm{pr}_{F=G}\{T(F_m, G_n) \leqslant K\} = 1-\alpha. \tag{3}$$

Then, simultaneously for all x,

$$\begin{aligned} 1-\alpha &= \mathrm{pr}_{F=G}\{T(F_m, G_n) \leqslant K\} = \mathrm{pr}_{F,G}\{T(F_m, G_{\Delta,n}) \leqslant K\} \\ &= \mathrm{pr}_{F,G}[h_*\{F_m(x)\} \leqslant G_{\Delta,n}(x) \leqslant h^*\{F_m(x)\}] \\ &= \mathrm{pr}_{F,G}[h_*\{F_m(x)\} \leqslant G_n\{\Delta(x)+x\} \leqslant h^*\{F_m(x)\}] \\ &= \mathrm{pr}_{F,G}(G_n^{-1}[h_*\{F_m(x)\}]-x \leqslant \Delta(x) < G_n^{-I}[h^*\{F_m(x)\}]-x). \end{aligned}$$

Thus we have the following theorem.

THEOREM 1. *If* (2) *and* (3) *hold, then*

$$[G_n^{-1}[h_*\{F_m(x)\}]-x,\ G_n^{-I}[h^*\{F_m(x)\}]-x) \tag{4}$$

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 (Pearson & Hartley, 1972, Table 55) so that $\mathrm{pr}_{F=G}(D_N \leqslant K_{S,\alpha}) = 1-\alpha$.

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

$$[G_n^{-1}\{F_m(x)-K_{S,\alpha}/M^{\frac{1}{2}}\}-x,\ G_n^{-I}\{F_m(x)+K_{S,\alpha}/M^{\frac{1}{2}}\}-x).$$

This band, which we call the S band and denote by $[S_*(x), S^*(x))$, was obtained by Switzer (1976) and by G.L. Sievers and should replace a similar band given by Doksum (1974).

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

$$[\Delta_*(x), \Delta^*(x)) = \left[Y\left(\left\langle nh_*\left(\frac{i}{m}\right)\right\rangle\right)-x,\ Y\left(\left[nh^*\left(\frac{i}{m}\right)\right]+1\right)-x\right) \tag{5}$$

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,

$$[S_*(x), S^*(x)) = \left[Y\left\{\left\langle n\left(\frac{i}{m}-K_{S,\alpha}/M^{\frac{1}{2}}\right)\right\rangle\right\}-x,\ Y\left\{\left[n\left(\frac{i}{m}+K_{S,\alpha}/M^{\frac{1}{2}}\right)\right]+1\right\}-x\right),$$

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

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

$$W_N = W_N(F_m, G_n) = M^{\frac{1}{2}} \sup_{\{x: a \leqslant F_m(x) \leqslant b\}} \frac{|F_m(x)-G_n(x)|}{\Psi\{H_N(x)\}}, \tag{6}$$

where $H_N(x) = \lambda F_m(x) + (1-\lambda)\, G_n(x)$, $\lambda = m/N$ and $0 \leqslant a < b \leqslant 1$. However if we choose $\Psi(t) = \{t(1-t)\}^{\frac{1}{2}}$, then we give approximately equal weight to each x in the sense that

$$M^{\frac{1}{2}}\{F_m(x)-G_n(x)\}/\Psi\{H_N(x)\}$$

has asymptotic variance independent of x. If we consider one-sided test statistics, without the absolute value, in the class (6) with $0 < a < b < 1$, this choice of Ψ asymptotically maximizes the minimum power when testing $H_0: F = G$ against $H_1: F(x) - G(x) \geqslant \delta$ for some x ($\delta > 0$) (Borokov & Sycheva, 1968).

To apply Theorem 1, we need to solve the inequality $|W_N(F_m, G_n)| \leqslant K$ for G_n. When $\Psi(t) = \{t(1-t)\}^{\frac{1}{2}}$ this inequality becomes

$$\{G_n(x) - F_m(x)\}^2 \leqslant K^2\{\lambda F_m(x) + (1-\lambda)\, G_n(x)\}\,[1 - \{\lambda F_m(x) + (1-\lambda)\, G_n(x)\}]/M$$

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

Let $c = K^2/M$, $u = F_m(x)$ and $v = G_n(x)$; then the inequality can be written as $d(v) \leqslant 0$, where

$$d(v) = \{1 + c(1-\lambda)^2\}\, v^2 - \{2u - c(1-\lambda)\,(2\lambda u - 1)\}\, v + u^2 - c\lambda u + c\lambda^2 u^2.$$

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

$$h^{\pm}(u) = \frac{u + \frac{1}{2}c(1-\lambda)\,(1-2\lambda u) \pm \frac{1}{2}\{c^2(1-\lambda)^2 + 4cu(1-u)\}^{\frac{1}{2}}}{1 + c(1-\lambda)^2}. \tag{7}$$

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

$$h^-\{F_m(x)\} \leqslant G_n(x) \leqslant h^+\{F_m(x)\}$$

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

Applying Theorem 1, we have shown the following.

Remark 2. Let $\mathrm{pr}_{F=G}(W_N \leqslant K) = 1-\alpha$; then the level $(1-\alpha)$ simultaneous confidence band for $\Delta(x)$ based on W_N with $\psi(t) = \{t(1-t)\}^{\frac{1}{2}}$ is

$$[G_n^{-1}[h^-\{F_m(x)\}] - x, \quad G_n^{-1}[h^+\{F_m(x)\}] - x), \quad x \in \{x : a \leqslant F_m(x) \leqslant b\}.$$

We refer to this band as the W band and write it $[W_*(x), W^*(x))$. 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 $a = 1-b = 0$. Figure 1 gives this band for X-samples and Y-samples from $N(0,1)$ and $N(1,4)$ distributions, respectively. Here $m = n = 100$, $\alpha = 0{\cdot}05$, and $K = 3{\cdot}02$ is obtained from Canner.

For $a > 0, b < 1$, asymptotic critical values K are given by Borokov & Sycheva (1968).

A third band can be obtained by considering the Renyi statistic

$$R_N = M^{\frac{1}{2}} \sup_{x \in \{x : F_m(x) \geqslant c\}} \frac{|F_m(x) - G_n(x)|}{H_N(x)}.$$

This statistic is reasonable 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 straightforward.

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

$$h_R^{\pm}(u) = \left\{\frac{1 \pm \lambda r/M^{\frac{1}{2}}}{1 \mp (1-\lambda)\, r/M^{\frac{1}{2}}}\right\} u,$$

then the R band $[R_*(x), R^*(x))$ is obtained by substituting $h_R^{\pm}$ for h_* and h^* in (5); x is required to be in $\{x : F_m(x) \geqslant c\}$.

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

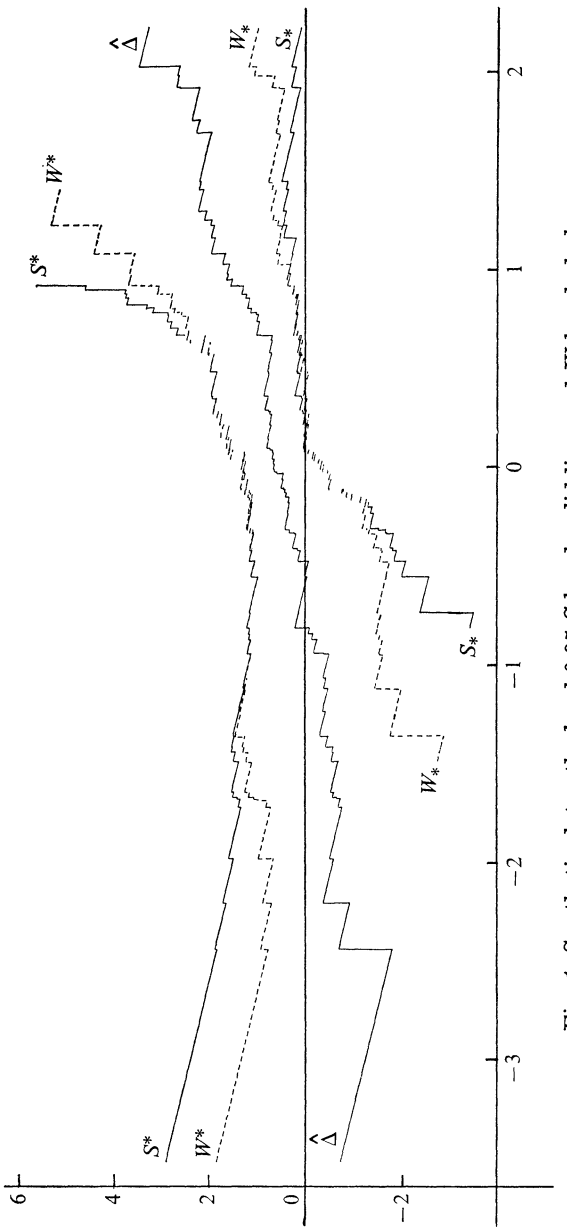

Fig. 1. Synthetic data; the level 0·95 S band, solid line, and W band, dashes.

3. COMPARISON OF THE NONPARAMETRIC BANDS

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

$$w_{M,\alpha}(x) = M^{\frac{1}{2}}G_n^{-I}[h^*\{F_m(x)\}] - M^{\frac{1}{2}}G_n^{-1}[h_*\{F_m(z)\}], \quad w_\alpha(x) = \lim_{M\to\infty} w_{M,\alpha}(x),$$

where the limit here and below is in probability. 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 $H_*(\epsilon, u)$ for $h_*(u)$ and $H^*(\epsilon, u)$ for $h^*(u)$. Now

$$\begin{aligned} w_{M,\alpha} &= \epsilon^{-1}[G^{-I}\{H^*(\epsilon,u)\} - G^{-I}(u)] - \epsilon^{-1}[G^{-1}\{H_*(\epsilon,u)\} - G^{-1}(u)] \\ &\quad + \epsilon^{-1}[G_n^{-I}\{H^*(\epsilon,u)\} - G^{-I}\{H^*(\epsilon,u)\}] - \epsilon^{-1}[G_n^{-1}\{H_*(\epsilon,u)\} - G^{-1}\{H_*(\epsilon,u)\}] \\ &\quad + \epsilon^{-1}\{G^{-I}(u) - G^{-1}(u)\} \\ &= \sum_{i=1}^{5} I_i, \end{aligned}$$

say.

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 the arguments of Doksum (1974, pp. 272–3) and a random change of time argument (Billingsley, 1968, pp. 144–6), we find that $I_3 + I_4$ converges in law to

$$\lambda^{\frac{1}{2}}[U\{F(x)\}/g\{\Delta(x)+x\} - U\{F(x)\}/g\{\Delta(x)+x\}]$$

on every interval $[F^{-1}(\delta), F^{-1}(1-\delta)]$ $(0 < \delta < \frac{1}{2})$, where U denotes the Brownian Bridge on $D[0, 1]$, and $\lambda = \lim (m/N)$ $(0 < \lambda < 1)$. It follows that $I_3 + I_4$ converges in probability to zero. We add the assumption that $H_*(\epsilon, u)$ and $H^*(\epsilon, u)$ converge to u as $\epsilon \to 0+$, and that $H_*(\epsilon, u)$ and $H^*(\epsilon, u)$ have right-hand derivatives $h_*(0, u)$ and $h^*(0, u)$ with respect to ϵ at $\epsilon = 0$ that are continuous in u.

Then we can compute the limits of I_1 and I_2 as follows:

$$\lim_{\epsilon\to 0+} I_1 = \lim_{\epsilon\to 0+} \frac{G^{-I}\{H^*(\epsilon,u)\} - G^{-I}\{H_*(0,u)\}}{\epsilon} = \frac{h^*\{0,F(x)\}}{g[G^{-1}\{F(x)\}]}, \quad \lim_{\epsilon\to 0+} I_2 = \frac{h_*\{0,F(x)\}}{g[G^{-1}\{F(x)\}]}.$$

To summarize, we have Theorem 2.

THEOREM 2. *Under the above stated conditions, the asymptotic widths of the bands* (4) *are given by*

$$w_\alpha(x) = \frac{h^*\{0,F(x)\} - h_*\{0,F(x)\}}{g[G^{-1}\{F(x)\}]}.$$

Let the asymptotic relative efficiency of two bands be the square of the ratio of the reciprocals of their asymptotic widths. Thus for two bands based on functions $\{h_*(u), h^*(u)\}$ and $\{k_*(u), k^*(u)\}$, we have that the efficiency is

$$e_{h,k}(x) = \left[\frac{k^*\{0,F(x)\} - k_*\{0,F(x)\}}{h^*\{0,F(x)\} - h_*\{0,F(x)\}}\right]^2.$$

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, $H_*(\epsilon, u) = u - K_{S,\alpha}\epsilon$ and $H^*(\epsilon, u) = u + K_{S,\alpha}\epsilon$; 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 X band, we replace c by $K^2\epsilon^2$ in (7) and differentiate with respect to ϵ to obtain $h_*(0,u)$ and $h^*(0,u)$. This yields

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

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

$$e_{S,W}(x_q) = (K_{W,\alpha}/K_{S,\alpha})^2 q(1-q) \quad (a \leqslant q \leqslant b),$$

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

Using the results of Borokov & Sycheva, we obtain the approximation to the asymptotic $K_{W,\alpha}$ given in Table 1. Combining this with the asymptotic tables for D_N, we obtain the efficiencies given in Table 2. The values for $q > 0{\cdot}5$ 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 centre of the X distribution at the expense of being worse in the tails up to $\pm x_a$. Technically, the asymptotic efficiency is infinite 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 $\alpha = 0{\cdot}01$, $K_{S,\alpha}/\sqrt{M} \geqslant 0{\cdot}1$ for $n \leqslant 531$, while for $\alpha = 0{\cdot}1$, $K_{S,\alpha}/\sqrt{M} \geqslant 0{\cdot}1$ for $n \leqslant 297$.

Table 1. *Asymptotic critical values of* W_N; $b = 1-a$

a \ α	0·2	0·1	0·04	0·02
0·25	2·109	2·482	2·879	3·138
0·1	2·482	2·789	3·138	3·371

Table 2. *The asymptotic efficiency* $e_{S,W}(x_q)$ *of the S band to the W band when* $\alpha = 0{\cdot}1$

(*a*) $a = 1-b = 0{\cdot}25$

α \ q	0·25	0·30	0·35	0·40	0·45	0·50
0·1	0·78	0·87	0·94	1·00	1·02	1·04

(*b*) $a = 1-b = 0{\cdot}1$

α \ q	0·1	0·2	0·3	0·4	0·5
0·1	0·47	0·83	1·09	1·25	1·30

Table 3. *Approximate values of the asymptotic efficiency,* $a = 1-b \simeq 0$

a \ q	0·01	0·05	0·1	0·2	0·3	0·4	0·5
0·05	0·06	0·26	0·50	0·89	1·17	1·34	1·39
0·01	0·05	0·24	0·46	0·82	1·08	1·24	1·29

Canner (1975) gives Monte Carlo critical values of W_N with $a = 1-b = 0$ when $n = m = 2000$ ($\alpha = 0{\cdot}05$) and $n = m = 1000$ ($\alpha = 0{\cdot}01$). Using these in place of asymptotic critical values, we obtain the approximations in Table 3. These values are close to the values of Table 2 with $a = 0{\cdot}1$.

We also computed the efficiencies for finite sample sizes, i.e. reciprocal ratios of widths for actual samples, using Canner's critical values and computer generated samples. Some results are given in Table 4 for X samples from a $N(0,1)$ distribution and Y samples from a $N(1,\sigma_2^2)$ distribution. These results are qualitatively close to the asymptotic results but

favour 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 centre 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 shift model while the S band does not; see (iii), § 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.

Table 4. *Finite sample size efficiency of the S band to the W band*; $\alpha = 0{\cdot}05$

(*a*) $m = n = 50$

σ_2 \ q	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8
0·5	∞/∞	0	0	0·61	1·01	1·00	0·76	∞/∞
2·0	∞/∞	0	0	0·61	1·18	1·00	0·66	∞/∞

(*b*) $m = n = 100$

σ_2 \ q	0·05	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9
0·5	∞/∞	0	0	0·87	0·97	1·18	1·18	0·80	0·65	∞/∞
2·0	∞/∞	0	0	0·80	0·90	1·14	1·19	1·09	0·56	∞/∞

For the R band, Theorem 2 yields the following 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. Using the results of Renyi (1953), we obtain the asymptotic $K_{R,\alpha}$ of Table 5. Combining this with the corresponding entries of Table 1, we compute the efficiencies of Table 6. The W band is preferable to the R band except when only small x_q are of interest.

Table 5. *Asymptotic critical values of* R_N

α \ c	0·2	0·1	0·05
0·1	3·921	5·881	8·545
0·05	4·484	6·726	9·772

Table 6. *Asymptotic efficiency of the R band to the W band*; $\alpha = 0{\cdot}1, a = 1-b$

a	c \ q	0·05	0·1	0·2	0·25	0·3	0·4	0·5	0·6	0·7
0·1	0·05	∞	0·96	0·43	0·32	0·25	0·16	0·11	0·07	0·05
	0·1	∞	2·02	0·90	0·67	0·52	0·34	0·22	0·15	0·10
0·25	0·2	∞	∞	∞	1·20	0·93	0·60	0·40	0·27	0·17

4. THE LOCATION-SCALE MODEL

In this section, we consider the model where

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

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

$$\Delta(x) = \mu_2 + \sigma_2 (x - \mu_1)/\sigma_1 - x.$$

In this model it is common to treat $\mu_2 - \mu_1$ as the parameter of interest. However, as seen in § 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

$$\text{pr}\{T_* \leqslant \Delta(x_1) \leqslant T^*,\ S_* \leqslant \Delta(x_2) \leqslant S^*\} = 1 - \alpha,$$

then for $x \in [x_1, x_2]$, the upper boundary would consist of the line connecting the points (x_1, T^*) and (x_2, S^*), while the lower boundary would be the line connecting (x_1, T_*) and (x_2, S_*). For $x > x_2$, the upper boundary would be the line through the points (x_1, T_*) and (x_2, S^*), and so on. Here x_1 and x_2 may be random; in the following, they are order statistics.

Remark 3. Suppose that a location-scale model holds. Let r_k, i_k and s_k $(k = 1, 2)$ be integers such that

$$\text{pr}_{F=G}\{Y(r_k) < X(i_k) < Y(s_k)\ \ (k = 1, 2)\} = 1 - \alpha. \tag{8}$$

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

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

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

Alternatively, Switzer (1976) 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 be the number of $Y_j \leqslant X(i_1)$, and Z_2 the number of $Y_j > X(i_2)$. Then the probability in (8) equals

$$\text{pr}_{F=G}(r_1 \leqslant Z_1 \leqslant s_1 - 1,\ n - s_2 + 1 \leqslant Z_2 \leqslant n - r_2),$$

and can be approximated using Lemma 1.

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*

$$V_k = M^{\frac{1}{2}}(Z_k/m - \xi_k)\{\beta_k(1 - \beta_k)\}^{-\frac{1}{2}}.$$

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

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 Theorem 2 of Sethuraman (1961) or a bivariate version of Hájek & Šidák (1967, problem 6, p. 195) gives the result.

The following remark gives a particular application of the lemma.

Remark 4. Let $i_1 = [m\beta] + 1$ and $i_2 = [m(1 - \beta)] + 1$ for some $\beta \in (0, \frac{1}{2})$. Then an asymptotic $(1 - \alpha)$ simultaneous confidence band for $\Delta(x)$ is determined by (9) with

$$(r_1 - \tfrac{1}{2})/n = \beta - c_\alpha\{\beta(1 - \beta)/M\}^{\frac{1}{2}}, \quad (s_1 - \tfrac{1}{2})/n = \beta + c_\alpha\{\beta(1 - \beta)/M\}^{\frac{1}{2}}, \tag{10}$$

for $r_2 = n + 1 - s_1$ and $s_2 = n + 1 - r_1$, where c_α satisfies $\text{pr}\{|V_k| \leqslant c_\alpha\ (k = 1, 2)\} = 1 - \alpha$ 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

$$L_M(x)\sqrt{M} \to \frac{2c_\alpha\{(\beta(1-\beta)\}^{\frac{1}{2}}\sigma_2}{(z_{1-\beta}-z_\beta)}\left\{\frac{|z_p-z_\beta|}{h(z_{1-\beta})}+\frac{|z_q-z_{1-\beta}|}{h(z_\beta)}\right\} = L(x),$$

say, for $x = \sigma_1 z_p + \mu_1$ with $\beta < p < 1-\beta$ and z_p the pth quantile of H. The convergence is in probability.

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

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

We call the band determined by (9) and (10) the O band, the order statistics band, and compare it with the W band of §2. In the location-scale model, the asymptotic width of the W band is $2K_{W,\alpha}\{p(1-p)\}^{\frac{1}{2}}\sigma_2/h(z_p)$. When H is normal, we obtain the asymptotic efficiencies given in Table 7. 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 too close to zero.

Table 7. *The asymptotic efficiency of the O band to the W band with* $a = 1-b = \beta$ *in the normal model* 0; $\alpha = 0{\cdot}1$

β \ p	0·10	0·20	0·25	0·30	0·40	0·50
0·10	2·05	1·43	1·30	1·22	1·13	1·10
0·25	—	—	1·64	1·53	1·42	1·38

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 $a = \sigma_2/\sigma_1$ and $b = \mu_2-\mu_1\sigma_2/\sigma_1$.

In the case that $m = n$, we will construct the likelihood ratio test for the hypothesis $H_0: a = a_0$, $b = b_0$ for fixed $a_0 \in (0, \infty)$ and $b_0 \in (-\infty, \infty)$. 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

$$L \propto \sigma_1^{-m}\sigma_2^{-n}\exp\left\{-\frac{1}{2\sigma_1^2}\sum_{i=1}^{m}(x_i-\mu_1)^2-\frac{1}{2\sigma_2^2}\sum_{j=1}^{n}(y_j-\mu_2)^2\right\}.$$

The maximum for unrestricted μ_1, μ_2, σ_1, σ_2 is well known. Substituting $\sigma_2^2 = a_0^2\sigma_1^2$ and $\mu_2 = b_0+a_0\mu_1$, the maximum of L under H_0 is found using standard methods. Let $(\bar{x}, \bar{y}, S_1^2, S_2^2)$ denote the usual unrestricted maximum likelihood estimate of $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$, then the likelihood ratio is

$$\Lambda_N = \sup_{H_0} L/\sup L = \frac{(a_0^2 S_1^2)^{\frac{1}{2}m}(S_2^2)^{\frac{1}{2}n}}{\{\lambda a_0^2 S_1^2+(1-\lambda)S_2^2+\lambda(1-\lambda)(\bar{y}-b_0-a_0\bar{x})^2\}^{\frac{1}{2}N}},$$

where $\lambda = m/N$.

The space of $(\mu_1, \mu_2, \sigma_1, \sigma_2)$ with $\sigma_2 = a_0\sigma_1$ and $\mu_2 = b_0+a_0\mu_1$ is linear with 2 dimensions. Thus, under H_0, by classical results, $-2\log\Lambda_N$ has a limiting chi-squared distribution with two degrees of freedom.

Let $x^2(1-\alpha)$ denote the $(1-\alpha)$th quantile of this distribution and let

$$K_\alpha = \exp\{x^2(1-\alpha)/N\};$$

then the acceptance region of the test when $m = n$ and $\lambda = \frac{1}{2}$ is

$$\{\tfrac{1}{2}a_0^2 S_1^2 + \tfrac{1}{2}S_2^2 + \tfrac{1}{4}(\bar{y} - b_0 - a_0\bar{x})^2\}/(a_0 S_1 S_2) \leqslant K_\alpha.$$

Write $b_0' = b_0 + a_0\bar{x}$; then the collection of (a_0, b_0') for which H_0 is accepted is the inside of an ellipse, namely

$$2(S_1 a_0 - K_\alpha S_2)^2 + (\bar{y} - b_0')^2 \leqslant 2S_2^2(K_\alpha^2 - 1). \tag{11}$$

Let $\mathscr{E}$ denote this ellipse and let

$$\delta^-(x) = \inf\{a_0 x + b_0 : (a_0, b_0') \in \mathscr{E}\} - x, \quad \delta^+(x) = \sup\{a_0 x + b_0 : (a_0, b_0') \in \mathscr{E}\} - x,$$

then $[\delta^-(x), \delta^+(x)]$ is the level $(1-\alpha)$ likelihood ratio confidence band for $\Delta(x)$. Theorem 3 specifies $\delta^\pm$.

THEOREM 3. *When $n = m$, the level $(1-\alpha)$ likelihood ratio confidence band for $\Delta(x)$ in normal models is given by*

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

Proof. We use the Cauchy–Schwarz inequality in the form

$$|\Sigma \alpha_i w_i| \leqslant (\Sigma \alpha_i^2 \Sigma w_i^2)^{\frac{1}{2}}, \tag{12}$$

with equality if and only if w_i is proportional to α_i. Simply note that (11) is equivalent to $w_1^2 + w_2^2 \leqslant d^2$, 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 5. The above proof is very similar to the derivation of Scheffé's simultaneous confidence intervals for contrasts. The interpretation is also similar: if the likelihood ratio test of $H_0 : \Delta(x) = 0$ for all x rejects, that is $a = 1$, $b = 0$, 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 $T_N(x) = M^{\frac{1}{2}}\{\hat{\Delta}(x) - \Delta(x)\}$. By using a Taylor expansion of $T_N(x)$ in terms of $\bar{X}$, $\bar{V}$, S_1 and S_2, we find that $T_N(x)$ converges in law to $T(x)$, say, equal to $\sigma_2(V_1 + tV_2/\sqrt{2})$, where $t = (x - \mu_1)/\sigma_1$ and V_1 and V_2 are independent standard normal variables. Write

$$\tau^2 = \operatorname{var}\{T(x)\} = \sigma_2^2(1 + \tfrac{1}{2}t^2), \quad \hat{\tau}^2 = S_2^2\{1 + \tfrac{1}{2}(x-\bar{x})^2/S_1^2\}.$$

Then we can use $\sup_x |T_N(x)|/\hat{\tau}$ as a pivot to obtain the following maximum likelihood band.

THEOREM 4. *An asymptotic $(1-\alpha)$ simultaneous confidence band for $\Delta(x)$ is given by*

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

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

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

$$\sup_t \{|(V_1 + tV_2/\sqrt{2})|/(1 + \tfrac{1}{2}t^2)^{-\frac{1}{2}}\} = \sup_t |l(t)|,$$

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 $(V_1^2 + V_2^2)^{\frac{1}{2}}$, which is the square root of a χ_2^2 variable. The result follows.

By standard asymptotic theory, the likelihood ratio and maximum likelihood bands should be asymptotically equivalent. This can be established directly when $n = m$ by using the first two terms in the Taylor expansion of e^z about $z = 0$ in the expression

$$K_\alpha = \exp\{x^2(1-\alpha)/N\}.$$

The likelihood ratio band is preferable since it is based on a more accurate approximation. The above expansion also yields the following.

Remark 6. The likelihood ratio and maximum likelihood bands both have asymptotic width $2\sigma_2 x(1-\alpha)(1+\frac{1}{2}t^2)^{\frac{1}{2}}$, 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 lose if in fact the correct model is normal.

We see that the asymptotic relative efficiency of the S band to the likelihood ratio band in normal models is

$$e_{S,L}(x_q) = \lim_{M\to\infty} \frac{(\text{width } L \text{ band})^2}{(\text{width } S \text{ band})^2} = \frac{\phi^2(t)\,x^2(1-\alpha)(1+\frac{1}{2}t^2)}{K_{S,\alpha}^2} = \frac{(1+\frac{1}{2}t^2)x_2^2(1-\alpha)}{2\pi e^{t^2} K_{S,\alpha}^2},$$

where $\phi(t)$ denotes the standard normal density. Similar expressions hold for the W and O bands. Some numerical results are given in Table 8. The efficiency of the S band is surprisingly low, much smaller than the familiar $2/\pi = 0{\cdot}64$.

Table 8. *Asymptotic efficiencies of the bands W, S and O with respect to the likelihood ratio band in normal models, $t = \Phi^{-1}(p)$, $\alpha = 0{\cdot}10$, $a = 1-b$*

Band	$p = 0{\cdot}1$	$p = 0{\cdot}2$	$p = 0{\cdot}3$	$p = 0{\cdot}4$	$p = 0{\cdot}5$
W, $a = 0{\cdot}1$	0·37	0·39	0·39	0·38	0·38
W, $a = 0{\cdot}25$	0	0	0·49	0·48	0·48
S	0·17	0·33	0·43	0·48	0·49
O, $\beta = 0{\cdot}1$	0·75	0·56	0·47	0·43	0·41
O, $\beta = 0{\cdot}25$	—	—	0·75	0·68	0·66

Table 9. *Finite sample size efficiency of the W band with $a = 1-b \doteq 0$ relative to the likelihood ratio band, $\alpha = 0{\cdot}05$ and $m = n$*

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

We also give in Table 9 some finite sample size 'efficiencies' computed for the same $N(0, 1)$ and $N(1, \sigma_2^2)$ samples as in § 3. 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 dis-

tributions will alter the level of the band. Also note that the O and likelihood ratio bands cannot 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.

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 are affected entirely differently by a tubercle bacillus 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, provided by Brian Tarkington, California Primate Research Center, University of California, Davis, are given in Table 10.

Table 10. *Weight gains of two groups of rats in grams*

Control	41·0	38·4	24·4	25·9	21·9	18·3	13·1	27·3	28·5	−16·9	26·0	17·4
Ozone	10·1	6·1	20·4	7·3	14·3	15·5	−9·9	6·8	28·2	17·9	−9·0	−12·9
Control	21·8	15·4	27·4	19·2	22·4	17·7	26·0	29·4	21·4	26·6	22·7	
Ozone	14·0	6·6	12·1	15·7	39·9	−15·9	54·6	−14·7	44·1	−9·0		

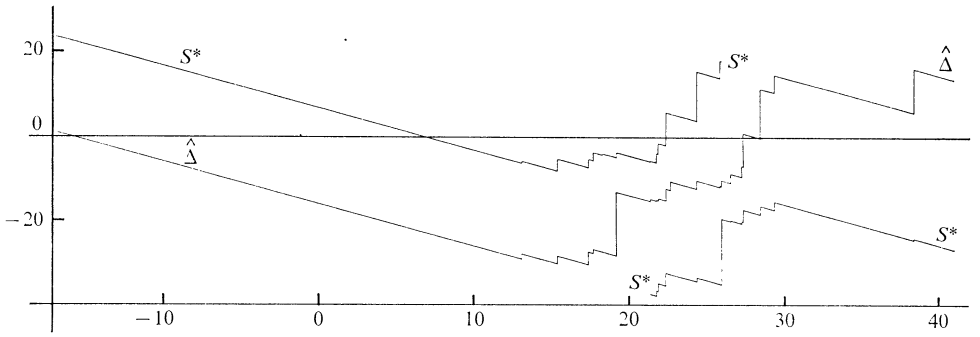

Fig. 2. The estimate $\hat{\Delta}(x)$ and the level 0·90 S band for the response function in the ozone experiment.

Figure 2 gives the estimate $\hat{\Delta}(x) = G_n^{-1}\{F_m(x)\} - x$ and the S-band with exact level 0·90. It is fairly obvious that $x = -16{\cdot}9$ is an outlier. If it is removed, only slight changes occur in the graphs beyond the removal of the long straight lines on the left of $\hat{\Delta}$ and S^*. Then $\hat{\Delta}$ suggests that, even though ozone reduces average weight gain, large weight gains are made even larger. Also S^* shows that weight gain is reduced significantly for x below the control weight gain 22·5.

From the S band, we cannot reject a shift model assumption, even though $\hat{\Delta}$ strongly indicates that it does not hold. This may be because the sample sizes are too small leaving the band too wide. The W band would be preferable, but we do not have the critical values for the sample sizes of this experiment.

Now $\hat{\Delta}$ with the outlier left out indicates that the response could well be linear and thus

the narrower O band could be used. For sample sizes $m = n = 22$, a good choice of β is 0·296 and then the O band is determined by

$$-37{\cdot}3 \leqslant \Delta(21{\cdot}4) \leqslant -7{\cdot}4, \quad -19{\cdot}3 \leqslant \Delta(26{\cdot}6) \leqslant 28{\cdot}0.$$

Thus, for this experiment, the S and O bands are remarkably close and the shift model assumption can still not be rejected.

We are indebted to P. Nanopoulos for programming the plots and the finite sample size efficiencies. The paper was prepared with the partial support of the National Science Foundation and the Norwegian Research Council. Doksum's work was partly done while visiting the Department of Mathematics, University of Oslo, and Sievers's work while visiting the Department of Statistics, University of California, Berkeley.

REFERENCES

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

BOROKOV, A. A. & SYCHEVA, N. M. (1968). On asymptotically optimal non-parametric criteria. *Theory Prob. Applic.* **13**, 359–93.

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**, 209–11.

DOKSUM, K. A. (1974). Empirical probability plots and statistical inference for nonlinear models in the two sample case. *Ann. Statist.* **2**, 267–77.

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

PEARSON, E. S. & HARTLEY, H. O. (1972). *Biometrika Tables for Statisticians*, Vol. 2. Cambridge University Press.

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

SETHURAMAN, J. (1961). Some limit theorems for joint distributions. *Sankhyā* A **23**, 379–86.

STECK, G. P., ZIMMER, W. J. & WILLIAMS, R. E. (1974). Estimation of parameters in acceleration models. In *Proc. Annual Reliability and Maintainability Symposium*. New York: I.E.E.E.

SWITZER, P. (1976). Confidence procedures for two sample problems. *Biometrika* **53**, 13–25.

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

[*Received February* 1976. *Revised May* 1976]