The Annals of Statistics
1974, Vol. 2, No. 2, 267-277

EMPIRICAL PROBABILITY PLOTS AND STATISTICAL INFERENCE FOR NONLINEAR MODELS IN THE TWO-SAMPLE CASE[1]

BY KJELL DOKSUM

University of California, Berkeley

Let X and Y be two random variables with continuous distribution functions F and G and means μ and ξ. In a linear model, the crucial property of the contrast $\Delta = \xi - \mu$ is that $X + \Delta =_{\mathscr{L}} Y$, where $=_{\mathscr{L}}$ denotes equality in law. When the linear model does not hold, there is no real number Δ such that $X + \Delta =_{\mathscr{L}} Y$. However, it is shown that if parameters are allowed to be function valued, there is essentially only one function $\Delta(\cdot)$ such that $X + \Delta(X) =_{\mathscr{L}} Y$, and this function can be defined by $\Delta(x) = G^{-1}(F(x)) - x$. The estimate $\hat{\Delta}_N(x) = G_n^{-1}(F_m(x)) - x$ of $\Delta(x)$ is considered, where G_n and F_m are the empirical distribution functions. Confidence bands based on this estimate are given and the asymptotic distribution of $\hat{\Delta}_N(\cdot)$ is derived. For general models in analysis of variance, contrasts that can be expressed as sums of differences of means can be replaced by sums of functions of the above kind.

1. Introduction. Consider the two-sample problem, which is the problem of comparing two populations on the basis of two independent samples $X_1, \cdots, X_m$ and $Y_1, \cdots, Y_n$, one from each population. Let F denote the distribution of the X_i and G the distribution of the Y_j. For this case, the assumption of a linear or shift model amounts to supposing that there exists a constant Δ, called the shift or translation parameter, such that

$$F(x) = G(x + \Delta) \quad \text{for all} \quad x . \tag{1.1}$$

When the linear model holds, then $X + \Delta$ has the same distribution as Y, where X is distributed according to F and Y according to G. In other words, the X_i when shifted the amount Δ, have the same distribution as the Y_j. Any comparison to be made between the two populations then depends on the parameter Δ, and much statistical theory concerns itself with this parameter. In general, if the linear model assumption is not satisfied, there is a function $\Delta(\cdot)$ such that

$$X + \Delta(X) \quad \text{has the same distribution as} \quad Y , \tag{1.2}$$

provided only that F is continuous. We write $X + \Delta(X) =_{\mathscr{L}} Y$ for (1.2) and call $\Delta(\cdot)$ a *shift function* since the X_i when shifted the amount $\Delta(X_i)$ have the same distribution as the Y_j. When the linear model assumption is not satisfied, $\Delta(\cdot)$ can be used as a measure of the difference between the two populations.

Received December 1972; revised May 1973.

[1] This research was supported in part by the National Science Foundation under GP-33697X.

AMS 1970 *subject classifications*. Primary 62G05, 62G15; Secondary 62G10, 62P10.

Key words and phrases. Nonlinear models, two-sample problem, shift function, empirical probability plot.

A function $\Delta(\cdot)$ satisfying (1.2) can be found in Lehmann's book (1974, Section 2.2). See Section 5.

Unfortunately, $X + \Delta(X) =_{\mathscr{L}} Y$ does not in general define $\Delta(\cdot)$. For instance, if X is normal $N(\mu, \sigma^2)$ and Y is $N(\mu + \xi, \sigma^2)$, then $\Delta_1(x) = \xi$ and $\Delta_2(x) = -2x + \xi + 2\mu$ both satisfy (1.2). However, if we defined $\Delta(x)$ as the "horizontal distance" between $F(x)$ and G, then $\Delta(\cdot)$ is well defined and satisfies (1.2). Thus we let $\Delta(x)$ be the smallest function satisfying (see Figure 1).

$$F(x) = G(x + \Delta(x)) . \tag{1.3}$$

It follows that $\Delta(x)$ can be expressed as $\Delta(x) = G^{-1}(F(x)) - x$, and that $\hat{\Delta}_N(x) = G_m^{-1}(F_m(x)) - x$ is a natural estimate of $\Delta(x)$, where F_m and G_n are the empirical distribution functions.

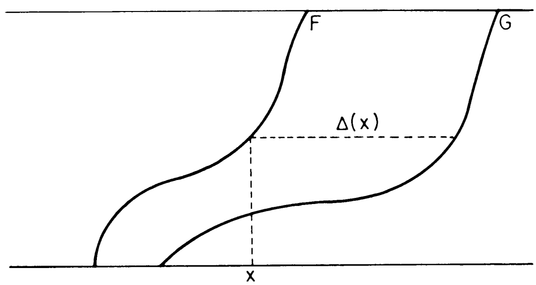

FIG. 1. The horizontal distance $\Delta(x)$ at x.

In Section 2, conditions are given under which $\Delta(x) = G^{-1}(F(x)) - x$ is the only function satisfying $X + \Delta(X) =_{\mathscr{L}} Y$; and $\Delta(\cdot)$ is compared with a function $\theta(\cdot)$ satisfying $Y - \theta(Y) =_{\mathscr{L}} X$. Moreover, it is shown that the following three conditions are equivalent:

(i) The linear model holds, (ii) $\theta(\cdot) = \Delta(\cdot)$, and (iii) $\Delta(x) \equiv$ constant.

The estimate $\hat{\Delta}_N(\cdot)$, which is an empirical probability plot minus the identity function, is considered in Section 3, where confidence bands for $\Delta(x)$ based on $\hat{\Delta}_N(x)$ are given. In Section 4, it is shown that the process $N^{\frac{1}{2}}[\hat{\Delta}_N(x) - \Delta(x)]$ converges weakly to a Gaussian process.

Finally, in Section 5, the important special case where the X's are control responses and the Y's are treatment responses is considered, and as an illustration, $\hat{\Delta}_N(x)$ and a confidence band for $\Delta(x)$ are computed using data from [3].

2. Properties of the shift function. Let $S(F) = \{x : 0 < F(x) < 1\}$ be the support of F.

DEFINITION 2.1. The shift function $\Delta(\cdot)$ for F and G is

$$\Delta(x) = \inf\{\Delta : F(x) \leqq G(x + \Delta)\} , \qquad x \in S(F) . \tag{2.1}$$

Let $G^{-1}(u) = \inf\{t : G(t) \geqq u, t \in \bar{S}(G)\}$, $u \in [0, 1]$, where $\bar{S}(G)$ denotes the closure of $S(G)$, then

$$\Delta(x) = G^{-1}(F(x)) - x , \qquad x \in S(F) . \tag{2.2}$$

When F is continuous, $G^{-1}(F(X))$ has the same distribution as Y, and therefore

$X + \Delta(X)$ has the same distribution as Y. If $\Delta(x)$ is to be a useful measure of differences between populations, we must show that it is essentially the only function satisfying $X + \Delta(X) =_{\mathscr{L}} Y$.

THEOREM 2.1. *If F is continuous, if $\Delta^*(\cdot)$ is any function such that $X + \Delta^*(X) =_{\mathscr{L}} Y$ and $\Delta^*(x) + x$ is non-decreasing for* a.a. $x(F)$, *then $\Delta^*(x) = \Delta(x)$ for* a.a $x(F)$.

PROOF. Let $h(x) = \Delta^*(x) + x$, then $h(x)$ and $G^{-1}(F(x))$ are two a.s. non-decreasing functions satisfying $h(X) =_{\mathscr{L}} G^{-1}(F(X))$. Thus h and $G^{-1}(F)$ are a.s. equal. Since $\Delta^*(x) = h(x) - x$ and $\Delta(x) = G^{-1}(F(x)) - x$, the result follows.

The next approach to the uniqueness problem is axiomatic. We give two desirable properties that a measure of shift between two populations should have and show that $\Delta(\cdot)$ is the only function satisfying these two properties. Let $\Delta^*(\cdot \mid F, G)$ denote a candidate for measure of shift between F and G, then Δ^* should satisfy

$$\text{(A.1)} \qquad \Delta^*(x \mid F, F) = 0\,, \quad x \in S(F)\,.$$

Let $\mu_Y(x) = \Delta^*(x \mid F, G) + x$, then $\mu_Y(x)$ is a measure of the location of Y corresponding to the measurement x on the first population (see Section 5). If Y is transformed by the increasing function h, then we require that $\mu_Y(\cdot)$ be transformed in the same way:

$$\text{(A.2)} \qquad \mu_{h(Y)}(x) = h(\mu_Y(x)) \quad \text{for} \quad x \in S(F) \quad \text{and all} \quad h \quad \text{continuous and increasing on} \quad S(G)\,.$$

PROPOSITION 2.1. *Suppose F and G are continuous and increasing on $S(F)$ and $S(G)$ respectively. If $\Delta^*(\cdot \mid F, G)$ is a function satisfying* (A.1) *and* (A.2), *then it must equal $\Delta(x \mid F, G) = G^{-1}(F(x)) - x$ on $S(F)$.*

PROOF. Set $h = F^{-1}G$ on $S(G)$, then by (A.1), $\Delta^*(x \mid F, Gh^{-1}) = \Delta^*(x \mid F, F) = 0$. But using (A.2), $\Delta^*(x \mid F, Gh^{-1}) = F^{-1}G(\Delta^*(x \mid F, G) + x) - x$. These two equations yield the result.

$\Delta(\cdot)$ has a natural competitor which is a function $\theta(\cdot)$ satisfying $Y - \theta(Y) =_{\mathscr{L}} X$. $\theta(y)$ can be defined as the horizontal distance from $G(y)$ to F and is given by $\theta(y) = y - F^{-1}(G(y))$. $\Delta(\cdot)$ and $\theta(\cdot)$ are not in general equal, but one can obtain one from the other. For instance, $\Delta(\cdot)$ is the inverse of the function $[t - \theta(t)]$ minus the identity function. Moreover, $\theta(t)$ and $\Delta(t)$ both measure horizontal distance between F and G, but at different points:

PROPOSITION 2.2. *Let $S^*(F) = \{t \in S(F) \text{ with } F^{-1}(F(t)) = t\}$, suppose that F and G are continuous, then $\Delta(t) = \theta(t + \Delta(t))$ for $t \in S^*(F)$ and $\theta(t) = \Delta(t - \theta(t))$ for $t \in S^*(G)$.*

PROOF. Since G is continuous, $G(t + \Delta(t)) = F(t)$. Combining this with $F^{-1}(F(t)) = t$ yields $-\theta(t + \Delta(t)) + [t + \Delta(t)] = t$ or $\Delta(t) = \theta(t + \Delta(t))$. $\theta(t) = \Delta(t - \theta(t))$ is similarly proved.

When does $\Delta(\cdot)$, in addition to $X + \Delta(X) =_{\mathscr{L}} Y$, have the property $Y - \Delta(Y) =_{\mathscr{L}} X$, or when is $\Delta(\cdot) = \theta(\cdot)$? It turns out that answering this question is equivalent to solving Euler's equation, and $\Delta(\cdot) = \theta(\cdot)$ only when the linear models hold:

PROPOSITION 2.3. *Suppose that F, G and $\Delta(\cdot)$ are continuous on the reals R. If $\Delta(t) = \theta(t)$ for $t \in R$, then there is a constant Δ such that $F(t) = G(t + \Delta)$ for all $t \in R$.*

PROOF. Since $\Delta(\cdot)$, $\theta(\cdot)$, F and G are continuous, F and G must be strictly increasing and $G^{-1}(F(t)) - t = t - F^{-1}(G(t))$. Set $H^{-1}(t) = G^{-1}(F(t))$ and $K(t) = H(t) - t$, then $H^{-1}(t) - t = t - H(t)$ or $H(H(t)) = 2H(t) - t$, and $K(t + K(t)) = H(H(t)) - (t + K(t)) = 2H(t) - t - (t + K(t)) = 2(K(t) + t) - t - (t + K(t)) = K(t)$. The equation $K(t + K(t)) = K(t)$ is called Euler's equation and is a special case of an equation solved by Nabeya (1972). Nabeya also points out that it has been considered by Kuratowski (1929) and Wagner (1959). The solution when K is continuous is given by $K(t) = \Delta$ for some constant Δ. It follows that $H(t) = t + \Delta$ and $F(t) = G(t + \Delta)$.

In order to test, on the basis of a confidence band for $\Delta(x)$ (see Section 3), whether the linear model assumption is satisfied, one needs the result that $\Delta(x)$ is a constant if and only if the linear model holds:

PROPOSITION 2.4. *If $F(x) = G(x + \Delta)$ for all x, then $P(\Delta(X) = \Delta) = P(\theta(Y) = \Delta) = 1$. If $\Delta(x) = \Delta$ for $x \in S(F)$, then $F(x) = G(x + \Delta)$ for all x. If F is strictly increasing on $S(F)$, then $F(x) = G(x + \Delta)$ for all x if and only if $\Delta(x) = \Delta$ for $x \in S(F)$.*

Next, we give some general properties of $\Delta(\cdot)$.

PROPOSITION 2.5. (i) *For arbitrary F and G, $X + \Delta(X)$ is stochastically no smaller than Y.* (ii) *The slope of the tangent to $\Delta(x)$ is bounded below by -1 in the sense that $[\Delta(x + \varepsilon) - \Delta(x)]\varepsilon^{-1} \geqq -1$ for all $\varepsilon > 0$, all $x \in S(F)$.* (iii) *When the expected values of X and Y exist, then $E(\Delta(X)) = E(Y) - E(X)$.*

Finally, $\Delta(x)$ and its estimate $\hat{\Delta}_N(x)$ have the following invariance properties. We write $\Delta(\cdot \mid F, G)$ for $\Delta(\cdot)$:

PROPOSITION 2.6. *Suppose h is continuous and increasing on $S(G)$, then $\Delta(x \mid F, Gh^{-1}) = h(\Delta(x \mid F, G) + x)$ for $x \in S(F)$. If h is also increasing and continuous on $S(G)$, then*

$$\Delta(h(x) \mid Fh^{-1}, Gh^{-1}) + h(x) = h(\Delta(x \mid F, G) + x) \qquad \textit{for} \quad x \in S(F)\,.$$

As an example, let F_a and G_a denote the distributions of $X + a$ and $Y + a$ respectively, then $\Delta(x \mid F, G_a) = \Delta(x \mid F, G) + a$ and $\Delta(x \mid F_a, G_a) = \Delta(x - a \mid F, G)$.

REMARK 2.1. The assumption that X and Y are independent is not needed in this section.

3. Empirical probability plots and distribution-free confidence bands for the shift function. To obtain an explicit expression for the empirical shift function

$\hat{\Delta}_N(x) = G_n^{-1}(F_m(x)) - x$, let $Y(1) \leqq \cdots \leqq Y(n)$ denote the ordered Y_j and let $\langle t \rangle$ denote the greatest integer less than t. For $x \in S(F_m)$, we have

$$\hat{\Delta}_N(x) = Y(\langle nF_m(x)\rangle + 1) - x\,, \tag{3.1}$$

which is the formula used in Figure 2. Rather than plotting this function, it would be just as useful to plot the points $(x(i), \hat{\Delta}_N(x(i))$, $i = 1, \cdots, m$, where $x(1) \leqq \cdots \leqq x(m)$ denote the ordered x_i. Since $\hat{\Delta}_N(x(i)) = G_n^{-1}(i/m) - x(i)$, we see that $(x(i), \hat{\Delta}_N(x(i)) + x(i))$, $i = 1, \cdots, m$, is an *empirical probability plot with* G_n taking the place of the usual parametric distribution function H in the probability plot $(x(i), H^{-1}(i/m))$, $i = 1, \cdots, m$. The present probability plot, rather than being a check of the distributional assumption H, is a check of the linear model assumption: If $(x(i), G_n^{-1}(i/m))$, $i = 1, \cdots, m$, fall near a straight line with slope one, then this supports the linear model assumption.

To get an idea of how reliable an estimate $\hat{\Delta}_N(t)$ is of $\Delta(t)$, we construct a confidence band based on $\hat{\Delta}_N(t)$ for $\Delta(t)$. This confidence band can then be turned into a test of the linear model assumption by checking if there is a horizontal line that falls in the band. Such a confidence band can easily be constructed from the one-sample Kolmogorov confidence bands for F and G. Let ε_1 and ε_2 be two numbers in $(0, \frac{1}{2})$ such that for continuous F and G

$$\begin{aligned} &P(F_m(x) - \varepsilon_1 \leqq F(x) \leqq F_m(x) + \varepsilon_1 \text{ for all } x) = 1 - \alpha_1\,, \\ &P(G_n(y) - \varepsilon_2 \leqq G(y) \leqq G_n(y) + \varepsilon_2 \text{ for all } y) = 1 - \alpha_2\,. \end{aligned} \tag{3.2}$$

By convention, $G_n^{-1}(t) = -\infty$ for $t < 0$ and $G_n^{-1}(t) = \infty$ for $t > 1$.

THEOREM 3.1. $(G_n^{-1}(F_m(x) - \varepsilon_1 - \varepsilon_2) - x, G_n^{-1}(F_m(x) + \varepsilon_1 + \varepsilon_2) - x)$ *is a simultaneous confidence band for* $\Delta(x)$ *with confidence coefficient at least* $(1-\alpha_1)(1-\alpha_2)$, i.e.

$$P(G_n^{-1}(F_m(x) - \varepsilon_1 - \varepsilon_2) - x \leqq \Delta(x) \leqq G_n^{-1}(F_m(x) + \varepsilon_1 + \varepsilon_2) - x \quad \textit{for all } x) \geqq (1 - \alpha_1)(1 - \alpha_2)\,.$$

PROOF. With probability $(1 - \alpha_2)$, $G_n(y) - \varepsilon_2 \leqq G(y) \leqq G_n(y) + \varepsilon_2$ for all y. It follows that $G_n^{-1}(u - \varepsilon_2) \leqq G^{-1}(u) \leqq G_n^{-1}(u + \varepsilon_2)$ for all u with probability $(1-\alpha_2)$. Similarly, with probability $(1 - \alpha_1)$, $F_m(x) - \varepsilon_1 \leqq F(x) \leqq F_m(x) + \varepsilon_1$ for all x. Since F_m and G_n are independent, the probability that both inequalities hold is $(1 - \alpha_1)(1 - \alpha_2)$, and the result follows.

When $F_m(x) + \varepsilon_1 + \varepsilon_2 > 1$, the only information the confidence band yields is $\Delta(x) \geqq G_n^{-1}(F_m(x) - \varepsilon_1 - \varepsilon_2)$, and similarly, when $F_m(x) - \varepsilon_1 - \varepsilon_2 < 0$, the confidence band reduces to an upper confidence boundary.

Other distribution-free confidence bands can be obtained if instead of the Kolmogorov confidence bands we use confidence bands obtained from statistics of the form

$$\sup_x |F(x) - F_m(x)|/q(F(x)) \qquad \text{or} \qquad \sup_x |F(x) - F_m(x)|/q(F_m(x))$$

for appropriate functions q. See for instance Steck (1971) or Birnbaum and Lientz (1969).

REMARK 3.1. If one wanted to check whether a log-linear or scale model $F(x) = G(ax)$, $a > 0$, is satisfied, one would check whether the empirical probability plot falls close to any line through the origin, since in this model $\Delta(x) = (a - 1)x$.

REMARK 3.2. It is not necessary to assume that F and G are continuous in Theorem 3.1, since when F and G are discrete, the confidence bands (3.2) are conservative.

4. Asymptotic theory. $\hat{\Delta}_N(x)$ is a consistent estimate of $\Delta(x)$, in fact $N^{\frac{1}{2}}(\hat{\Delta}_N(x) - \Delta(x))$ converges weakly to a Gaussian process. This result, which is established in this section, is a start towards developing asymptotic methods for $\Delta(x)$ (x fixed), $\sup_x \Delta(x)$, and $\inf_x \Delta(x)$. Before the result can be applied, one would have to investigate the effect of estimating the parameters in the asymptotic normal distribution of $N^{\frac{1}{2}}(\hat{\Delta}_N(x) - \Delta(x))$.

Define $\lambda_N = (m/N)$ and suppose that there is a constant $\lambda \in (0, 1)$ such that $\lambda_N \to \lambda$ as $N \to \infty$. Assume that the supports of F and G are two non-empty finite intervals (a, b) and (c, d) respectively. Extend the definition of $\hat{\Delta}_N(x)$ from $S(F_m) = [X(1), X(m))$ to $[a, b]$ by $\hat{\Delta}_N(x) = \inf\{t: G_n(t) \geqq F_m(x), t \in [c, d]\} - x$, and the definition of $\Delta(x)$ from $S(F) = (a, b)$ to $[a, b]$ by $\Delta(x) = G^{-1}(F(x)) - x$, $x \in [a, b]$. $N^{\frac{1}{2}}(\hat{\Delta}_N(x) - \Delta(x))$ is a member of the space D of functions on $[a, b]$ that are right continuous and have left-hand limits. On this space, we use the usual Skorohod topology (e.g. [1, page 111]).

$W_0(t)$ will denote a Brownian Bridge on [0, 1], that is, a Gaussian process with mean zero and covariance function $s(1 - t)$, $0 \leqq s \leqq t \leqq 1$.

THEOREM 4.1. *Suppose that $G(t)$ has a continuous derivative $g(t)$ satisfying $0 < g(t) < \infty$ on $[c, d]$. Then $N^{\frac{1}{2}}[\hat{\Delta}_N(x) - \Delta(x)]$ converges in distribution (weakly) to the Gaussian process*

$$[g(G^{-1}(F(x)))]^{-1}[\lambda(1 - \lambda)]^{-\frac{1}{2}} W_0(F(x)) \,. \tag{4.1}$$

PROOF. Let $U_1, \cdots, U_n$ be independent random variables with each U_i uniformly distributed over [0, 1] and $U_1, \cdots, U_n$ independent of $X_1, \cdots, X_m$. Let H_n denote the empirical distribution function for $U_1, \cdots, U_n$ and let $H_n^{-1}(u) = \inf\{t: H_n(t) \geqq u, t \in [0, 1]\}$, $u \in [0, 1]$. Since $u \leqq G(t)$ if and only if $G^{-1}(u) \leqq t$, $G^{-1}(U_1), \cdots, G^{-1}(U_n)$ has the same distribution as $Y_1, \cdots, Y_n$; and the process $G^{-1}(H_n^{-1}(u))$ has the same probability distribution as the process $G_n^{-1}(u)$. Moreover $G^{-1}(H_N^{-1})$ is independent of G_m, so the process

$$D_N(x) =_{\text{def}} N^{\frac{1}{2}}[G^{-1}(H_n^{-1}(F_m(x))) - G^{-1}(F(x))] \,, \qquad x \in [a, b] \,, \tag{4.2}$$

has the same probability distribution as $N^{\frac{1}{2}}(\hat{\Delta}_N(x) - \Delta(x))$, $x \in [a, b]$. Set $u = F(x)$ and $u_N = H_n^{-1}(F_m(x))$, then, since $u_N \neq u$ a.s.,

$$D_N(x) = \frac{G^{-1}(u_N) - G^{-1}(u)}{u_N - u} N^{\frac{1}{2}}(u_N - u) \quad \text{a.s.} \tag{4.3}$$

It will next be shown that $N^{\frac{1}{2}}(u_N - u)$ converges in distribution to $W_0(F(x))/[\lambda(1-\lambda)]^{\frac{1}{2}}$. $N^{\frac{1}{2}}(u_N - u) = N^{\frac{1}{2}}[H_n^{-1}(F_m(x)) - F_m(x)] + N^{\frac{1}{2}}[F_m(x) - F(x)]$. By a well-known result (e.g. [1, page 141]), $N^{\frac{1}{2}}[F_m(x) - F(x)] = (N/m)^{\frac{1}{2}} m^{\frac{1}{2}}[F_m(x) - F(x)]$ converges in distribution to $W(F(x))/\lambda^{\frac{1}{2}}$, where $W(t)$ is a Brownian Bridge on [0, 1]. Set $V_n(u) = n^{\frac{1}{2}}[H_n^{-1}(u) - u]$. It is known [9], [10] that the U_i can be constructed on a probability space Ω in conjunction with a Brownian Bridge V such that $\sup_{0 \leqq u \leqq 1} |V_n(u) - V(u)| \to_e 0$ where $\to_e$ denotes convergence everywhere on the probability space Ω. Note that

$$\sup_{a \leqq x \leqq b} |V_n(F_m(x)) - V(F(x))| \\ \leqq \sup_{0 \leqq u \leqq 1} |V_n(u) - V(u)| + \sup_{a \leqq x \leqq b} |V(F_m(x)) - V(F(x))| .$$

The first term on the right-hand side of the inequality converges to zero, as already remarked. By the Glivenko–Cantelli theorem, $\sup_x |F_m(x) - F(x)| \to 0$ a.s. Since V is a.s. uniformly continuous, the second term on the right-hand side converges a.s. to zero. It follows from this that $N^{\frac{1}{2}}[H_n^{-1}(F_m(x)) - F_m(x)] = (N/n)^{\frac{1}{2}} V_n(F_m(x))$ converges in law to $V(F(x))/(1-\lambda)^{\frac{1}{2}}$. By construction, V and W are independent, and $(W(F(x))/\lambda^{\frac{1}{2}}) + V(F(x))/(1-\lambda)^{\frac{1}{2}}$ has the same law as $W_0(F(x))/[\lambda(1-\lambda)]^{\frac{1}{2}}$. $\operatorname{Sup}_{0 \leqq u \leqq 1} |H_n^{-1}(u) - u| = \sup_{0 \leqq u \leqq 1} |H_n(u) - u|$ and two applications of the Glivenko–Cantelli theorem imply that $u_N = H_n^{-1}(F_m(x))$ satisfies $\sup_x |u_N - u| \to 0$ a.s., where $u = F(x)$. Next we establish that $A_N(u) =_{\text{def}} [G^{-1}(u_N) - G^{-1}(u)]/(u_N - u)$ converges uniformly to $1/g(G^{-1}(u)) =_{\text{def}} A'(u)$. Since A' is continuous on [0, 1], it is uniformly continuous. Thus given $\varepsilon > 0$, there exists $\delta > 0$ such that $\sup_{|t-s| \leqq \delta} |A'(t) - A'(s)| < \varepsilon$. By the mean value theorem, $|A_N(u) - A'(u)| = |A'(t) - A'(u)|$ for some $t \in (u, u_N)$. Since $\sup_x |u_N - u| \to 0$ a.s., then for a.a. ω there exists $N_{0,\omega}$ such that $|u - u_N| < \delta$ whenever $N > N_{0,\omega}$. Thus for a.a. ω and $N > N_{0,\omega}$,

$$\sup_{0 \leqq u \leqq 1} |A_N(u) - A'(u)| \leqq \sup_{|t-s| \leqq \delta} |A'(t) - A'(s)| < \varepsilon .$$

The result follows from this.

REMARK 4.1. (a) If convergence of $N^{\frac{1}{2}}(\hat{\Delta}_N(x) - \Delta(x))$ to a normal variable is only needed for fixed x, the assumption of Theorem 4.1 can be relaxed to "$G(t)$ has a derivative $g(t)$ at $t = G^{-1}(F(x))$ satisfying $0 < g(t) < \infty$." (b) The assumption that the supports of F and G are finite intervals can be omitted if one restricts the definition of $N^{\frac{1}{2}}(\hat{\Delta}_N(x) - \Delta(x))$ to a finite closed interval.

5. An example involving control vs. treatment. When the X's are control responses and the Y's are treatment responses, the following model has been considered by Lehmann (1974, Section 2.2): "Suppose that there exists a function $\delta(\cdot)$ such that the treatment adds the amount $\delta(x)$ when the response of the untreated subject is x. Then the distribution of the treatment response Y is that of the random variable $X + \delta(X)$." Thus if we assume that $\delta(x) + x$ is nondecreasing, then according to Theorem 2.1, $\delta(\cdot)$ equals the shift function $\Delta(\cdot)$. When the above model holds, $\delta(\cdot) = \Delta(\cdot)$ will be called the *treatment effect* *function*. Note that it is a rather special model since it assumes that two subjects with the same control response will react the same way to the treatment. However, it is a much better approximation to the real situation than the linear model which essentially assumes that all subjects react the same way to the treatment regardless of their control response.

As an illustration, $\hat{\Delta}_y(x)$ was computed for the survival times in days of the guinea pigs used in Bjerkedal's (1960) study of the effect of tubercle bacilli. The two samples considered are the first two columns of Bjerkedal's Table 6 which contain a control group (column one) of 107 guinea pigs, and a treatment group (column two) of 61 guinea pigs which received a dose of tubercle bacilli (see Table 1 of the present paper). The relevant linear model parameter is the difference of the means. Its estimate, the difference of the sample means, reveals that the mean life of the treated animals is significantly shorter than the mean of the untreated group. $\hat{\Delta}_y(\cdot)$ and its confidence band yields much more information about the two populations (see Figure 2). The confidence band shows first of all that the linear model assumption is not satisfied since a horizontal line can not be drawn in the band. About the effect of the treatment, $\hat{\Delta}_y(\cdot)$ indicates that the weak members of the population benefit from the treatment, while the treatment is extremely harmful to the strong members of the population. Furthermore, the stronger the member of the population, the more harmful the treatment. Here, an animal in the control group is called weak if he is prone to die at an early age, and strong if he is prone to die at an advanced age. The upper boundary of the confidence band shows that the treatment is significantly harmful only for guinea pigs with control response greater than 250 days. The present confidence band does not allow one to conclude that the treatment is significantly beneficial for weak members of the population.

The above example illustrates some general principles that lead one to believe that the linear model assumption is often not satisfied (not even approximately):

TABLE 1

Survival time in days up to 735 days

Control group x				18	36	50	52	86	87	89	91	102
Treatment group y				76	93	97	107	108	113	114	119	136
x	108	114	114	115	118	119	120	149	160	165	166	167
y	137	138	139	152	154	154	160	164	164	166	168	178
x	167	173	178	189	209	212	216	273	278	279	292	341
y	179	181	181	183	185	194	198	212	213	216	220	225
x	355	367	380	382	421	421	432	446	455	463	474	505
y	225	244	253	256	259	265	268	268	270	283	289	291
x	545	546	569	576	590	603	607	608	621	634	634	637
y	311	315	326	326	361	373	373	376	397	398	406	459
x	638	641	650	663	685	688	725	735				
y	466	592	598									

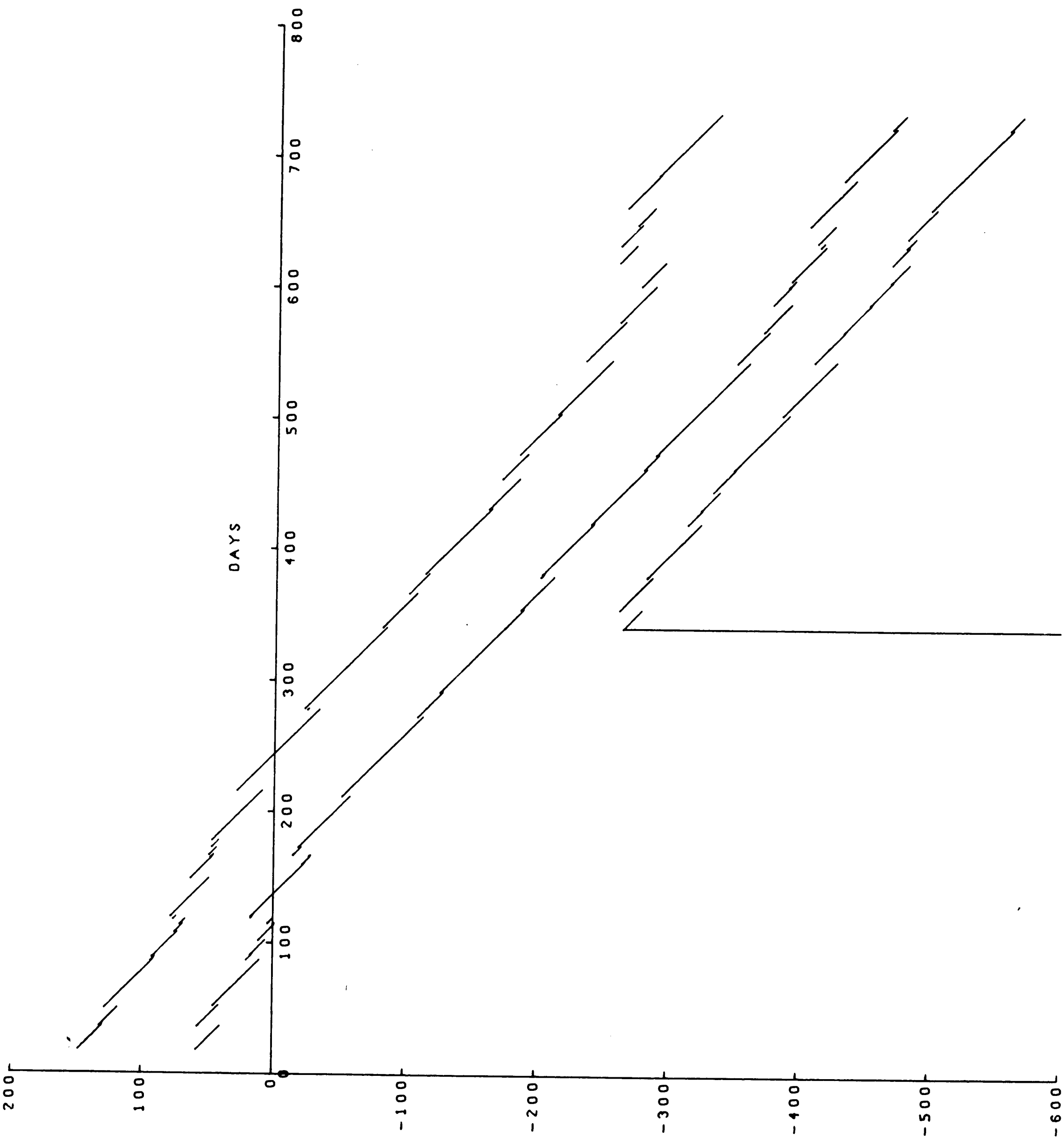

FIG. 2. The empirical shift function and confidence band with confidence coefficient .90.

Frequently, the members of a population can be said to have a certain property such as "prone to die at an advanced age", "prone to grow tall", "prone to get high blood pressure", "prone to learn fast", etc. For such populations, if we consider two members that have this property to a different degree, then we would expect a treatment designed to influence this property to affect these two members differently. In other words, there is interaction between the proneness property and the treatment. $\Delta(\cdot)$ is a measure of this interaction. For the experimental results in Table 1, Bjerkedal and Palmer (1959) call this interaction an interaction between experimental tuberculosis and extraneous forces. The usual way to treat this problem in analysis of variance is to use block designs. However, this is only possible for characteristics that can be measured before the experiment. Often, there are characteristics that cannot be measured before the experiment, such as the proneness to die early in the guinea pigs example. Thus blocking should be used to remove variation due to measurable characteristics, and then estimates based on shift functions should be used to measure the effect of the treatment and the interaction between the treatment and these unmeasurable characteristics or proneness properties.

REMARK 5.1. Looking at Figure 2, one sees that a model assuming that $\Delta(x)$ is a line $\alpha x + \beta$ could be a good fit in the example. This amounts to assuming a shift and scale model, i.e. $F(x) = G((\alpha + 1)x + \beta)$. A good estimate of $\Delta(x)$ would be the regression line through the points $\{(x(i), \hat{\Delta}_N(x(i))), i = 1, \cdots, m\}$. This estimate and more general models in analysis of variance will be treated in a forthcoming paper.

Acknowledgment. The author is grateful to Steinar Bjerve for programming the plots in Figure 2.

REFERENCES

[1] BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

[2] BIRNBAUM, Z. W. and LIENTZ, B. P. (1969). Exact distributions for some Rényi-type statistics. *Zastos. Mat.* **10** 179-192.

[3] BJERKEDAL, T. (1960). Acquisition of resistance in guinea pigs infected with different doses of virulent tubercle bacilli. *Amer. J. Hygiene* **72** 130-148.

[4] BJERKEDAL, T. and PALMER, C. E. (1959). Competing forces of mortality in long-term animal experiments. *Amer. J. Hygiene* **70** 246-253.

[5] KURATOWSKI, K. (1929). Sur une equation fonctionelle. *Sprawozdania z posiedzén Tow. Nauk. Warszawskiego* **22** 160-161.

[6] LEHMANN, E. L. (1974). *Statistical Methods Based on Ranks.* Holden Day, San Francisco.

[7] NABEYA, S. (1974). On the functional equation $f(p + qx + rf(x)) = a + bx + cf(x)$. *Aequationes Math. Univ. of Waterloo.*

[8] PYKE, R. and SHORACK, G. (1968). Weak convergence of a two-sample empirical process and a new approach to Chernoff-Savage theorems. *Ann. Math. Statist.* **39** 755-771.

[9] SHORACK, G. (1972). Functions of order statistics. *Ann. Math. Statist.* **43** 412-427.

[10] STECK, G. P. (1971). Rectangle probabilities for uniform order statistics and the probability that the empirical distribution function lies between two distribution functions. *Ann. Math. Statist.* **42** 1-11.

[11] WAGNER, R. (1959). Eidentige lösungen der funktionalgleichung $f[x + f(x)] = f(x)$. *Elem. Math.* **14** 73-78.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720