NON-PARAMETRIC THEORY: SCALE AND LOCATION PARAMETERS

D. A. S. FRASER

1. Summary. In §2 a result in measure theory is obtained. The remainder of this paper, §3 to §11, contains results in the branch of statistics called non-parametric theory; these results in part are based on the measure result of §2.

The measure result concerns a class of probability distributions—those distributions having a probability density function on the real line and for which a fraction p of the probability is on the negative axis and a fraction $q = 1 - p$ is on the positive axis. Corresponding to a sample of n the functional form is obtained for a statistic having expectation zero for all distributions in the class; such a statistic is referred to as an unbiased estimate of zero.

In §3 a reasonable definition of location parameter for the continuous distributions on the real line requires it to be the p-percentile, that is, the point having a total probability p to the left of it. In §4 confidence regions for this parameter are characterized, confidence bounds are shown to be based on order statistics, and confidence regions with certain optimum properties are obtained. In §5 several problems in hypothesis testing on the location parameter are considered and most powerful and most powerful unbiased tests are obtained. A bivariate analogue of one of these problems is considered in §6.

In §7 reasonable definitions are considered for scale and location-scale parameters for continuous distributions on the real line. For the scale parameter a result of negative nature is obtained in §8: that similar texts do not exist for the hypothesis that specifies a value for the scale parameter.

In §9 a formulation is given for non-parametric tolerance regions. A particular type of these, distribution-free upper tolerance bounds, was treated by Robbins in 1944. His condition, obtained under an assumption of continuity, is shown to be necessary but not sufficient in the general case; a bound chooses the order statistics with fixed but arbitrary probabilities.

In §10 some results in estimation theory obtained by Lehmann and Scheffé are extended to permit wider application in non-parametric theory. Two examples of estimation in non-parametric theory are considered in §11.

2. A measure problem with applications in statistics. Some results will be obtained for probability distributions over R^n. First we define some classes of measures on the real line R^1. Let $\mathfrak{F}$ be the class of probability measures on R^1, $\mathfrak{F}_0$ be the subclass of distributions absolutely continuous with respect to Lebesgue measure, $\mathfrak{F}_0(p)$ the subclass of $\mathfrak{F}_0$ whose elements have $F(0) = p$, $\mathfrak{F}_1$ be the class of discrete distributions with probability at a finite number

Received November 5, 1952; in revised form May 5, 1953.

of points, $\mathfrak{F}_2$ be the class of uniform distributions over a finite number of intervals, and $\mathfrak{F}_3$ be the class having a probability density of the form

$$c(\theta_1, \ldots, \theta_n) \exp \{-x^{2n} - \sum \theta_i x^i\}.$$

From these distributions we derive measures over R^n, the power product measure induced by a measure or distribution on R^1. Letting $F(x_1, \ldots, x_n)$ be the distribution function obtained from $F(x)$, then

$$F(x_1, \ldots, x_n) = \prod_1^n F(x_i).$$

We designate by $\mathfrak{F}_i^n$ the class of measures over R^n which is obtained from $\mathfrak{F}_i$:

$$\mathfrak{F}_i^n = \left\{ \prod_1^n F(x_i) | F(x) \in \mathfrak{F}_i \right\}.$$

To give an outline of some previous results concerning the classes $\mathfrak{F}_i^n$, we need the concept of a complete class of measures. Let $\mu_\theta(A)$ be a probability measure over a space $\mathfrak{X}$ with a σ-algebra of subsets $\mathfrak{A}$; that is, μ_θ satisfies

(1) $\mu_\theta(A) \geqslant 0,\ A \in \mathfrak{A}$.

(2) $\mu_\theta(\mathfrak{X}) = 1$.

(3) If $A_i \in \mathfrak{A}$ and $A_i \cap A_j = \phi (i \neq j)$, then

$$\mu_\theta\left(\bigcup_1^\infty A_i\right) = \sum_1^\infty \mu_\theta(A_i).$$

The class of measures $\{\mu_\theta(A) | \theta \in \Omega\}$ *is complete if* $\int f(x) d\mu_\theta(x) \equiv 0$ *for all* θ *implies* $f(x) = 0$ *almost everywhere* $\{\mu_\theta(A)\}$.

In non-parametric theory applied to distributions over R^n, the order statistics play an important role; we define a statistic $T(x_1, \ldots, x_n) = (x_{(1)}, \ldots, x_{(n)})$, the "order statistics," where $x_{(1)}, \ldots, x_{(n)}$ are the numbers $x_1, \ldots, x_n$ arranged in order of magnitude. Obviously any function of $x_1, \ldots, x_n$ which can be expressed as a function of $T(x_1, \ldots, x_n)$ is a symmetric function. Corresponding to any distribution over R^n the statistic $T(x_1, \ldots, x_n)$ will have an induced probability distribution.

In 1946, Halmos **(1)** showed that the distributions of $T(x_1, \ldots, x_n)$ corresponding to $\mathfrak{F}_1^n$ were complete. Lehmann in **(2)** proved a similar result for $\mathfrak{F}_3^n$. In **(3)** the author showed the same for $\mathfrak{F}_2^n$. The distributions of $T(x_1, \ldots, x_n)$ corresponding to $\mathfrak{F}_0(p)$ are, however, not complete (unless $p = 0, 1$); we prove here some results for $\mathfrak{F}_0(p)$ which are the natural extensions of the concept of completeness.

A statistic $\phi(x_1, \ldots x_n)$ *is an unbiased estimate of a real valued function* $g(F)$ *of the distributions* $F(x)$ *of a class* $\mathfrak{G}$ *if*

$$\int_{R^n} \phi(x_1, \ldots, x_n) \prod dF(x_i) = g(F)$$

for all distributions $F \in \mathfrak{G}$.

Using this definition we have the

https://doi.org/10.4153/CJM-1954-007-7 Published online by Cambridge University Press

THEOREM 2.1. *For the distributions $\mathfrak{F}_0(p)$ a necessary and sufficient condition that a function of $T(x_1, \ldots x_n)$ be an unbiased estimate of zero is that it have the form*

$$\phi(x_1, \ldots, x_n) = \sum_i \alpha(x_i)\psi(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) \tag{2.1}$$

almost everywhere where $\psi(x_1, \ldots x_{n-1})$ is an arbitrary bounded measurable symmetric function, and

$$\begin{aligned} \alpha(x) &= -q = -(1-p) && \text{if } x < 0, \\ &= +p && \text{if } x > 0. \end{aligned} \tag{2.2}$$

Note: The theorem gives the form of a function $\phi(x_1, \ldots, x_n)$ satisfying

$$\int_{R^n} \phi(x_1, \ldots, x_n) \prod dF(x_i) \equiv 0 \tag{2.3}$$

for all $F(x) \in \mathfrak{F}_0(p)$. If we relax our requirement of absolute continuity and consider the class $\mathfrak{F}(p)$ of all distributions on R^1 having $F(0) = p = F(0-0)$, then the only change is that (2.1) is required to hold everywhere. The proof is obtained by trivial changes in the lemmas.

Proof. We first note that $\phi(x_1, \ldots, x_n)$ is bounded almost everywhere. Otherwise there would exist a sequence of numbers $c_1, c_2, \ldots$ approaching ∞ and sets $S_1, S_2, \ldots$

$$S_i = \{(x_1, \ldots, x_n) \mid |\phi(x_1, \ldots, x_n)| \geqslant c_i\}$$

such that each has positive Lebesgue measure. For any such set it is possible to obtain a rectangular set which is more than $\frac{1}{2}$, say, filled (Lebesgue) with points of S_i. On the basis of the sequence of rectangular sets it is possible to define a density function for which $E\{|\phi|\}$ would not exist. The Theorem assumes that all expectations exist equal to zero; hence a contradiction.

The proof proper then obtains from the following three lemmas.

LEMMA 2.1. *If $\phi(x_1, \ldots, x_n)$ is a symmetric unbiased estimate of zero for $\mathfrak{F}_0(p)$, then almost everywhere $(z_1, \ldots z_n, y_1, \ldots, y_n) \in]-\infty, 0[^n \times]0, \infty[^n$*

$$\sum_{r=0}^{n} \sum_{x=z^{n-r}y^r} p^{n-r}q^r\phi(x_1, \ldots, x_n) = 0, \tag{2.4}$$

where the summation with subscript $x = z^{n-r}y^r$ is taken over the $\binom{n}{r}$ terms obtained by replacing r x's with y's and $n - r$ x's with z's.

Proof. In **(3)** a complete sufficient statistic was given for a sample of n from an arbitrary bivariate distribution over $]-\infty, 0[\times]0, \infty[$. Letting $(z_1, y_1), \ldots, (z_n, y_n)$ be the sample elements, the statistic is $\{(z_1, y_1), \ldots, (z_n, y_n)\}$. Then if $\phi(z_1, y_1, \ldots, z_n, y_n)$ is an unbiased estimate of zero for these distributions the symmetrized form of ϕ,

$$\bar{\phi}(z_1, y_1, \ldots, z_n, y_n) = \frac{1}{n!}\sum_P \phi(z_{i_1}, y_{i_1}, \ldots, z_{i_n}, y_{i_n})$$

https://doi.org/10.4153/CJM-1954-007-7 Published online by Cambridge University Press

where the summation is over all permutations $(i_1, \ldots, i_n)$ of $(1, \ldots n)$, will be zero almost everywhere.

Let $f(z, y)$ be an arbitrary probability density function over $]-\infty, 0[\times]0, \infty[$, and $f_-(z)$ and $f_+(y)$ be respectively the z and y marginal densities. If, outside the present range of definition of $f_-(z)$, $f_+(y)$ on the real line, we give them the value zero, then $g(x) = pf_-(x) + qf_+(x)$ is the density of a distribution belonging to $\mathfrak{F}_0(p)$.

Now if $\phi(x_1, \ldots, x_n)$ is a symmetric unbiased estimate of zero for $\mathfrak{F}_0(p)$, then for $g(x)$ defined above we have

$$\begin{aligned} 0 &= \int_{R^n} \phi(x_1, \ldots, x_n) \prod g(x_i) \prod dx_i \\ &= \sum_0^n \binom{n}{r} p^{n-r} q^r \int_0^\infty \cdots \int_0^\infty \int_{-\infty}^0 \cdots \int_{-\infty}^0 \phi(y_1, \ldots, y_r, z_{r+1}, \ldots, z_n) \\ &\qquad\qquad \prod_1^r f_+(y_i) \prod_{r+1}^n f_-(z_i) \prod_1^r dy_i \prod_{r+1}^n dz_i \\ &= \sum_0^n \binom{n}{r} p^{n-r} q^r \int_{S^n} \phi(y_1, \ldots, y_r, z_{r+1}, \ldots, z_n) \prod_1^n f(z_i, y_i) \prod_1^n dz_i \prod_1^n dy_i \\ &= \int_{S^n} \sum_0^n \binom{n}{r} p^{n-r} q^r \phi(y_1, \ldots, y_r, z_{r+1}, \ldots, z_n) \prod_1^n f(z_i, y_i) \prod_1^n dz_i \prod_1^n dy_i. \end{aligned}$$

But from **(3)** we have that the symmetrized form of the integrand is zero almost everywhere. This completes the proof.

LEMMA 2.2. *If for all* $(z_1, \ldots, z_n, y_1, \ldots, y_n) \in]-\infty, 0[^n \times]0, \infty[^n$,

$$\sum_{x=z,y} \phi(x_1, \ldots, x_n) = 0, \tag{2.5}$$

where the summation is over all terms obtained by replacing each x *by either* z *or* y, *then*

$$\phi(x_1, \ldots, x_n) = \sum \alpha'(x_i) \psi_i(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) \tag{2.6}$$

where

$$\begin{aligned} \alpha'(x) &= +1 && \text{if } x > 0, \\ &= -1 && \text{if } x < 0. \end{aligned}$$

Proof. The proof is obtained by induction. For $n = 1$ the lemma is obvious; assume it holds for $n - 1$. From (2.5) we have

$$\begin{aligned} \sum_{x=z,y} \phi(z_1, x_2, \ldots, x_n) &= \sum \phi(\overset{*}{z_1}, x_2, \ldots, x_n) \\ &= -\sum \phi(y_1, x_2, \ldots, x_n) \\ &= -\sum \phi(\overset{*}{y_1}, x_2, \ldots, x_n). \end{aligned}$$

Then

$$\sum_{x=z,y} [\phi(z_1, x_2, \ldots, x_n) + \phi(\overset{*}{y_1}, x_2, \ldots, x_n)] = 0,$$
$$\sum_{x=z,y} [\phi(y_1, x_2, \ldots, x_n) - \phi(\overset{*}{y_1}, x_2, \ldots, x_n)] = 0.$$

https://doi.org/10.4153/CJM-1954-007-7 Published online by Cambridge University Press

But by the inductive argument,

$$\phi(z_1, x_2, \ldots, x_n) + \phi(\overset{*}{y_1}, x_2, \ldots, x_n)$$
$$= \alpha'(x_2)\psi_2(z_1, x_3, \ldots, x_n) + \alpha'(x_3)\psi_3(z_1, x_2, x_4, \ldots, x_n) + \ldots,$$
$$\phi(y_1, x_2, \ldots, x_n) - \phi(\overset{*}{y_1}, x_2, \ldots, x_n)$$
$$= \alpha'(x_2)\psi_2(y_1, x_3, \ldots, x_n) + \alpha'(x_3)\psi_3(y_1, x_2, x_4, \ldots, x_n) + \ldots.$$

These two equations together imply (2.6).

If in addition $\phi(x_1, \ldots, x_n)$ is symmetric, then the ψ functions can be the same:

$$\psi_i(x_1, \ldots, x_{n-1}) = \psi(x_1, \ldots, x_{n-1}).$$

Also the symmetrized ψ function is uniquely determined. For suppose we have two determinations:

$$\phi(x_1, \ldots, x_n) = \sum \alpha'(x_i) \quad \psi(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$$
$$= \sum \alpha'(x_i) \ \overset{*}{\psi}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n).$$

By subtraction,

$$0 = \sum \alpha'(x_i) \ \xi(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n). \tag{2.7}$$

We now prove by induction that ξ is identically zero. For $n = 1$, the statement is obvious; assume it holds for $n - 1$. We have

$$-\alpha'(x_1) \ \xi(x_2, \ldots, x_n) = \alpha'(x_2) \ \xi(x_1, x_3, \ldots, x_n) + \ldots$$
$$+ \alpha'(x_n) \ \xi(x_1, x_2, \ldots, x_{n-1}),$$

and the left hand side is independent of x_1 ($x_1 > 0$ or $x_1 < 0$). The assumption for $n - 1$ implies that the right hand side term by term is independent of x_1 ($x_1 > 0$ or $x_1 < 0$). Also when x_1 changes sign, so does the left side and hence the right side term by term. Thus we have

$$\xi(x_1, \ldots, x_{n-1}) = \alpha'(x_1) \ \overset{*}{\xi}(x_2, \ldots, x_{n-1})$$

and from symmetry

$$\xi(x_1, \ldots, x_{n-1}) = \prod_1^{n-1} \alpha'(x_i) \ \overset{*}{\xi}.$$

Substitution in (2.7) then gives $\xi^* = 0$, $\xi(x_1, \ldots, x_n) = 0$, and therefore the uniqueness of $\psi(x_1, \ldots, x_{n-1})$ in the symmetric case.

LEMMA 2.3. *If* $\phi(x_1, \ldots, x_n)$ *is symmetric and satisfies* (2.4), *then* $\phi(x_1, \ldots, x_n)$ *has the form* (2.1).

Proof. Letting $i(x_1, \ldots, x_n)$ be the number of positive x's and defining $\phi'(x_1, \ldots, x_n)$ by

$$\phi'(x_1, \ldots, x_n) = p^{n - i(x_1, \ldots, x_n)} \ q^{i(x_1, \ldots, x_n)} \ \phi(x_1, \ldots, x_n),$$

then (2.4) becomes

https://doi.org/10.4153/CJM-1954-007-7 Published online by Cambridge University Press

$$\sum_{x=z,y} \phi\,(x_1, \ldots, x_n) = 0. \tag{2.8}$$

This last relation need only hold almost everywhere. However, if we replace $\phi(x_1, \ldots, x_n)$ by its average over a rectangular cube with sides $\delta_1, \ldots, \delta_n$ centred on $(x_1, \ldots, x_n)$, then (2.8) holds everywhere (points distant $n^{\frac{1}{2}}$ max δ_i from the diagonal planes excepted). Lemma 2.2 then gives

$$\phi'(x_1, \ldots, x_n) = \alpha'(x_1)\,\psi_1'(x_2, \ldots, x_n) + \ldots + \alpha'(x_n)\,\psi_n'(x_1, \ldots, x_{n-1}). \tag{2.9}$$

Since the ψ's are just linear combinations of values of ϕ's, then by the Radon-Nikodym theorem the above form for $\phi'(x_1, \ldots, x_n)$ holds almost everywhere as well as on the average as obtained. From the symmetry of ϕ we may have $\psi_i = \psi$ independent of i.

Substituting for ϕ' in terms of ϕ, using $\alpha(x)$, and appropriately defining ψ, (2.9) becomes (2.1).

3. Definition of a location parameter. In sampling from a probability distribution over the real line the statistician, assuming a non-parametric hypothesis, will usually envisage a class of distributions as general as $\mathfrak{F}_0$: $\mathfrak{F}_0$ is the class of distributions having a density function on the real line. For this class of distributions we consider what real parameters can legitimately be called location parameters.

By a real valued parameter for a class of distributions $\mathfrak{G}$ is meant a real number for each $F \in \mathfrak{G}$; that is, a real valued function $\xi(F)$ defined over $\mathfrak{G}$. A reasonable requirement for $\xi(F)$ to be called a location parameter might be given by the

DEFINITION. *$\xi(F)$ is a* location parameter *if, for any $F,G \in \mathfrak{G}$ for which $F(x) = G(f(x))$ where $f(x)$ is monotone nondecreasing,*

$$\xi(G) = f(\xi(F)).$$

The meaning of this definition is more apparent in terms of random variables. Let X be a random variable with distribution $F(x)$; then $Y = f(X)$ has distribution $G(y)$. The condition is that the value of the location parameter also be transformed by the function $f(x)$. Many location parameters in parametric problems satisfy this condition.

As an immediate consequence of this definition, we obtain that $\xi(F)$ is a percentile of the distribution and that there is a number p such that $F(\xi(F)) = p$ (with the obvious modification if F is not continuous). For if we assume that $\mathfrak{G}$ contains a continuous distribution G, then for any $F \in \mathfrak{G}$ there is a monotone non-decreasing function $f_F(x)$ such that $G(x) = F(f_F(x))$. The definition gives $\xi(F) = f_F(\xi(G))$. This uniquely determines $\xi(F)$ and we have

$$F(\xi(F)) = F(f_F(\xi(G))) = G(\xi(G)) = p,$$

where p is the constant $G(\xi(G))$.

Restricting ourselves to percentiles as the reasonable non-parametric location parameter, we define the p-percentile by

$$\xi_p(F) = F^{-1}(p).$$

This definition is not always unique. For if there is an interval over which $F(x)$ is constant at the value p, then the definition gives all the points of that interval. However this is not a real drawback.

4. Confidence regions for the location parameter. From the results of Theorem 2.1, it is possible to characterize similar β confidence regions for the location parameter $\xi_p(F)$.

Let $S(x_1, \ldots x_n; R)$ be a set on the real line for each $(x_1, \ldots, x_n; R)$ where $0 \leqslant R \leqslant 1$. Thus $S(x_1, \ldots, x_n; R)$ is a mapping from $R^n \times [0,1]$ into the space of subsets of R^1. We require that the characteristic function,

$$\Phi_\theta(x_1, \ldots, x_n; R) = 1 \qquad \text{if } \theta \in S(x_1, \ldots, x_n; R),$$

$$= 0 \qquad \text{if } \theta \notin S(x_1, \ldots, x_n; R),$$

should be measurable in $(x_1, \ldots, x_n; R)$ for each θ. We say that $S(x_1, \ldots, x_n; R)$ *is a β randomized confidence region for the parameter $\xi_p(F)$ if*

$$\mathrm{Pr}_F(\xi_p(F) \in S(X_1, \ldots, X_n; R)) \geqslant \beta$$

for all $F \in \mathfrak{F}_0$ where $(X_1, \ldots, X_n)$ is a sample of n from the distribution F and R is assumed to be uniformly distributed on $[0,1]$. Similarly $S(x_1, \ldots x_n; R)$ *is a similar β randomized confidence region for $\xi_p(F)$ if the inequality with β is replaced by equality.*

We investigate similar β confidence regions. For a confidence region $S(x_1, \ldots, x_n; R)$, it is possible to define a characteristic function by eliminating the dependence on R:

$$\phi_\theta(x_1, \ldots, x_n) = \frac{1}{n!}\sum \mathrm{Pr}_R\{\theta \in S(x_{i_1}, \ldots, x_{i_n}; R)\}$$

where the summation is over the $n!$ permutations $(i_1, \ldots, i_n)$ of $(1, \ldots, n)$. It is easily seen that ϕ_θ is the conditional expectation of Φ_θ given values for $x_{(1)}, \ldots, x_{(n)}$ and that it is symmetric in $x_1, \ldots, x_n$.

If $S(x_1, \ldots, x_n; R)$ is a similar β confidence region for $\xi_p(F)$ then

$$E_F(\phi_\theta(X_1, \ldots, X_n)) \equiv \beta$$

for all $F \in \mathfrak{F}_0$ having $\xi_p(F) = \theta$. Thus we obtain

$$E_F(\phi_\theta(X_1, \ldots, X_n) - \beta) = 0$$

when $\xi_p(F) = \theta$.

For simplicity consider the case $\theta = 0$; then the condition on ϕ is that $\phi_0(x_1, \ldots, x_n) - \beta$ be an unbiased estimate of zero for $\mathfrak{F}_0(p)$. By Theorem 2.1 we have

https://doi.org/10.4153/CJM-1954-007-7 Published online by Cambridge University Press

$$\phi_0(x_1, \ldots, x_n) - \beta = \sum \alpha(x_i)\, \psi(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) \quad \text{a.e.}$$

and similarly

$$\phi_\theta(x_1, \ldots, x_n) - \beta = \sum \alpha(x_i - \theta)\, \psi_\theta(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) \quad \text{a.e.}$$

Thus in part we have

THEOREM 4.1. *A necessary and sufficient condition that $S(x_1, \ldots, x_n; R)$ be a similar β confidence region for $\xi_p(F)$ is that*

$$\begin{aligned}(4.1)\quad \phi_\theta(x_1, \ldots, x_n) - \beta \\ &= p\psi_\theta(x_{(2)}, \ldots, x_{(n)}) + \ldots + p\psi_\theta(x_{(1)}, \ldots, x_{(n-1)}) && \text{if } \theta \leqslant x_{(1)} \\ &= -q\psi_\theta(x_{(2)}, \ldots, x_{(n)}) - \ldots - q\psi_\theta(x_{(1)}, \ldots, x_{(i-1)}, x_{(i+1)}, \ldots, x_{(n)}) \\ &\quad + p\psi_\theta(x_{(1)}, \ldots, x_{(i)}, x_{(i+2)}, \ldots, x_{(n)}) + \ldots + p\psi_\theta(x_{(1)}, \ldots, x_{(n-1)}) \\ & && \text{if } x_{(i)} < \theta \leqslant x_{(i+1)} \\ &= -q\psi_\theta(x_{(2)}, \ldots, x_{(n)}) - \ldots - q\psi_\theta(x_{(1)}, \ldots, x_{(n-1)}) && \text{if } x_{(n)} < \theta.\end{aligned}$$

Proof. The necessity is proved above; the sufficiency follows immediately from Theorem 2.1.

It is interesting to note that for any $\phi_\theta(x_1, \ldots, x_n)$ of the form (4.1) with values restricted to [0,1] there exists a confidence region with $\phi_\theta(x_1, \ldots, x_n)$ as characteristic function. Let

$$S(x_1, \ldots, x_n; R) = \{\theta \mid \phi_\theta(x_1, \ldots, x_n) \geqslant R\};$$

then it is easily seen that $S(x_1, \ldots x_n; R)$ is a similar β confidence region for $\xi_p(F)$.

The above theorem allows us to determine the form of upper confidence bounds, and the next theorem shows that such bounds are necessarily a choice of the order statistics.

THEOREM 4.2. *If $S(x_1, \ldots, x_n; R) =]-\infty, u(x_1, \ldots, x_n; R)[$ is a similar confidence region for $\xi_p(F)$, then*

$$\begin{aligned}(4.2)\quad \bar{u}(x_1, \ldots, x_n; R) &= -\infty && \textit{with probability } p_0(x_1, \ldots, x_n) \\ &= x_{(1)} && \textit{with probability } p_1(x_1, \ldots, x_n) \\ &\ldots && \ldots \\ &= x_{(n)} && \textit{with probability } p_n(x_1, \ldots, x_n) \\ &= +\infty && \textit{with probability } p_{n+1}(x_1, \ldots, x_n)\end{aligned}$$

where $\bar{u}(x_1, \ldots, x_n; R)$ is symmetric in $x_1, \ldots, x_n$ and is obtained from $u(x_1, \ldots, x_n; R)$ by incorporating into R the randomization of the $n!$ permutations $(i_1, \ldots, i_n)$ of $(1, \ldots, n)$, and where $\sum p_i = 1$, $p_i > 0$, and there is a set of functions $\{P_0(x_1, \ldots, x_{n-1}), \ldots, P_{n-1}(x_1, \ldots, x_{n-1})\}$ such that

https://doi.org/10.4153/CJM-1954-007-7 Published online by Cambridge University Press

$$p_i(x_1, \ldots, x_n) = \sum_0^{i-1} P_j(x_{(1)}, \ldots, x_{(i-1)}, x_{(i+1)}, \ldots, x_{(n)})$$
$$+ p \sum_{j=1}^{n-i} P_i(x_{(1)}, \ldots, x_{(i)}, \ldots, x_{(i+j-1)}, x_{(i+j+1)}, \ldots, x_{(n)})$$
$$- q \sum_{j=1}^{i-1} P_{i-1}(x_{(1)}, \ldots, x_{(j-1)}, x_{(j+1)}, \ldots, x_{(i)}, \ldots, x_{(n)}),$$

and conversely.

Proof. The randomization inherent in the ordering of the x_i's can without loss of generality be combined with that provided by the random element R; thus we assume that $S(x_1, \ldots, x_n; R)$ and $u(x_1, \ldots x_n; R)$ are symmetric in the x's.

By examining its definition we see that the characteristic function of an upper confidence bound, $\phi_\theta(x_1, \ldots, x_n)$, is a monotone non-increasing function of θ. From Theorem 4.1, $\phi_\theta(x_1, \ldots, x_n) - \beta = \bar{\phi}_\theta(x_1, \ldots, x_n)$ has almost everywhere the form

(4.3) $\bar{\phi}_\theta(x_1, \ldots, x_n)$

$$= p\psi_\theta(x_{(2)}, \ldots, x_{(n)}) + \ldots + p\psi_\theta(x_{(1)}, \ldots, x_{(n-1)}) \qquad \text{if } \theta \leqslant x_{(1)}$$
$$= -q\psi_\theta(x_{(2)}, \ldots, x_{(n)}) - \ldots - q\psi_\theta(x_{(1)}, \ldots, x_{(i-1)}, x_{(i+1)}, \ldots, x_{(n)})$$
$$+ p\psi_\theta(x_{(1)}, \ldots, x_{(i)}, x_{(i+2)}, \ldots, x_{(n)}) + \ldots + p\psi_\theta(x_{(1)}, \ldots, x_{(n-1)})$$
$$\text{if } x_{(1)} < \theta \leqslant x_{(i+1)}$$
$$= -q\psi_\theta(x_{(2)}, \ldots, x_{(n)}) - \ldots - q\psi_\theta(x_{(1)}, \ldots, x_{(n-1)}) \qquad \text{if } x_{(n)} < \theta.$$

Consider two values of θ, say θ_1, θ_2 $(\theta_1 < \theta_2)$, and $2n$ values of x, say $z_1, \ldots, z_n, y_1, \ldots, y_n$ $(z_1 < \ldots < z_n < \theta_1 < \theta_2 < y_1 < \ldots < y_n)$ and assume that any ϕ using the θ's and x's from these sets is of the form (4.3). (The average over any small cube centred over the particular points in R^n will always satisfy (4.3).) Writing

$$\phi^*(x_1, \ldots, x_n) = \bar{\phi}_{\theta_1}(x_1, \ldots, x_n) - \bar{\phi}_{\theta_2}(x_1, \ldots, x_n),$$
$$\psi^*(x_2, \ldots, x_n) = \psi_{\theta_1}(x_2, \ldots, x_n) - \psi_{\theta_2}(x_2, \ldots, x_n)$$

we have from the monotonicity of ϕ_θ that $\phi^*(x_1, \ldots, x_n) \geqslant 0$. Then for r z's, say $z_1, \ldots, z_r$, and $n - r$ y's, say $y_{r+1}, \ldots, y_n$, we have

(4.4) $$p\psi^*(z_1, \ldots, z_r, y_{r+1}, \ldots, y_{n-1}) + \ldots + p\psi^*(z_1, \ldots, z_r, y_{r+2}, \ldots, y_n)$$
$$- q\psi^*(z_1, \ldots, z_{r-1}, y_{r+1}, \ldots, y_n) - \ldots - q\psi^*(z_2, \ldots, z_r, y_{r+1}, \ldots, y_n) \geqslant 0.$$

Thus, for any set of n numbers chosen from $z_1, \ldots, z_n, y_1, \ldots, y_n$ we write down the n values of the ψ^* function by deleting successively one of the n numbers. To each of these we attach the coefficient $+p$ or $-q$ according as the deleted number is a y or a z. From the hypotheses it follows that the algebraic sum is always nonnegative. Thus there are $\binom{2n}{n}$ inequalities and $\binom{2n}{n-1}$ ψ^* values.

https://doi.org/10.4153/CJM-1954-007-7 Published online by Cambridge University Press

We proceed to show that the ϕ^* values are all zero. Consider $\binom{2n}{n-1}$ vectors each with $\binom{2n}{n}$ coordinates. For each set of $n-1$ numbers chosen from the $2n$, we define a vector with a coordinate corresponding to each set of n numbers which can be chosen from the $2n$; the value of this coordinate is zero if the $n-1$ numbers giving the vector are not included in the n numbers giving the coordinate, and is $+p$, $-q$ if the $n-1$ vector numbers are included in the n coordinate numbers and if the additional number is a y, z. The inequalities (4.4) then say that a linear combination with weights ψ^* of these vectors gives a vector in or on the boundry of the first orthant. We wish to show that the only such combination is a combination with zero coefficients and this of course gives the zero vector (zero ϕ^* values).

We now define a vector which has coordinates $c(x_1, \ldots, x_n)$ all positive and is orthogonal to each of the $\binom{2n}{n-1}$ vectors defined above. This orthogonality condition for the vector corresponding to $(z_1, \ldots, z_r, y_{r+2}, \ldots, y_n)$ is

$$pc(z_1, \ldots, z_r, y_{r+1}, \ldots, y_n) + \ldots + pc(z_1, \ldots, z_r, y_1, y_{r+2}, \ldots, y_n)$$
$$-qc(z_1, \ldots, z_r, z_n, y_{r+2}, \ldots, y_n) - \ldots - qc(z_1, \ldots, z_{r+1}, y_{r+2}, \ldots, y_n) = 0.$$

Defining $c(x_1, \ldots, x_n)$ by

$$c(x_1, \ldots, x_n) = \binom{n}{i(x_1, \ldots, x_n)}^{-1} p^{n-i(x_1,\ldots,x_n)} q^{i(x_1,\ldots,x_n)},$$

the above equation becomes

$$p(r+1)\binom{n}{r}^{-1} p^r q^{n-r} - q(n-r)\binom{n}{r+1}^{-1} p^{r+1} q^{n-r-1} = 0,$$

which is obviously true.

Thus each of the $\binom{2n}{n-1}$ vectors and hence any linear combination thereof will be vectors in the linear subspace perpendicular to the $c(x_1, \ldots, x_n)$ vector and passing through the origin. Because each coordinate of the $c(x_1, \ldots, x_n)$ vector is positive, each vector in the linear subspace must have at least one coordinate negative unless all are zero. Thus the only vector in the subspace and in the first orthant is the zero vector. This means that the $\binom{2n}{n}$ values of $\phi^*(x_1, \ldots, x_n)$ are zero.

From $\phi^*(x_1, \ldots, x_n) = 0$, we obtain that

$$\bar{\phi}_{\theta_1}(x_1, \ldots, x_n) = \bar{\phi}_{\theta_2}(x_1, \ldots, x_n)$$

so long as all x values lie outside $[\theta_1, \theta_2]$. This equality of the $\bar{\phi}$ functions implies by the concluding remarks to Lemma 2.2 that the ψ functions are also the same

$$\psi_{\theta_1}(x_1, \ldots, x_{n-1}) = \psi_{\theta_2}(x_1, \ldots, x_{n-1})$$

and hence $\psi_\theta(x_1, \ldots, x_{n-1})$ is constant valued as a function of θ except possibly for jumps at the points $x_1, \ldots, x_{n-1}$. Letting P_i stand for the jump at $x_{(i)}$, we have

$$(4.5) \qquad \psi_\theta(x_1, \ldots, x_{n-1}) = \sum_{j=0}^{i-1} P_j(x_1, \ldots, x_{n-1}) \qquad \text{if } x_{(i-1)} < \theta \leqslant x_{(i)}.$$

For fixed $x_1, \ldots, x_n$, $\bar{u}(x_1, \ldots, x_n; R)$ is a real valued random variable; it has distribution function

$$\Pr_R\{\bar{u}(x_1, \ldots, x_n; R) \geqslant \theta\} = \Pr_R\{\theta \in S(x_1, \ldots, x_n; R)\} = \phi_\theta(x_1, \ldots, x_n).$$

Thus from (4.5) and (4.3), we obtain (4.2). The converse follows from (4.3) and the definitions of the functions involved. This completes the proof.

It might seem at first sight that a result similar to that above would apply to confidence intervals, viz., that they would be the interval between two order statistics chosen randomly. We give two examples of confidences intervals for which the bounds are not both order statistics.

Example 1. Let $\beta = .25$, $p = .5$, and $n = 2$. Let $f(x_1,x_2)$ be any real valued function such that $x_{(1)} \leqslant f(x_1,x_2) \leqslant x_{(2)}$. Then a .25 similar confidence interval for the median is

$$\begin{aligned} S(x_1, x_2; R) &= [x_{(1)}, f(x_1, x_2)] & \text{if } R < .5\,, \\ &= [f(x_1, x_2), x_{(2)}] & \text{if } R \geqslant .5\,. \end{aligned}$$

Example 2. Let $\beta = \frac{4}{9}$, $p = \frac{1}{3}$, and $n = 2$. Then a $\frac{4}{9}$ similar confidence interval for the p-percentile is

$$\begin{aligned} S(x_1, x_2; R) &= [x_{(1)}, x_{(2)}] & \text{if } x_{(2)} \leqslant 0, \\ &= [x_{(1)}, 0] & \text{if } x_{(1)} \leqslant 0 < x_{(2)}, \\ &= [0, x_{(2)}] & \text{if } 0 < x_{(1)}. \end{aligned}$$

These are easily checked. A high confidence level can be obtained for either example by taking a larger value of n.

The theorems above supply us with some indication of the form of similar confidence regions for the location parameter $\xi_p(F)$. It is perhaps natural then to look for confidence regions possessing certain optimum properties. The following properties which one might require of confidence regions were introduced by Wald. However, since there is almost a complete analogy between confidence region theory and hypothesis testing theory we shall use the names which are standard for tests. We have the following definitions:

DEFINITION 4.1. *The* power function *of a confidence region for* $\xi_p(F)$ *is*

$$\text{(4.6)} \quad P_F(\theta) = \Pr_F(\theta \notin S(X_1, \ldots, X_n; R)) = E_F(1 - \phi_\theta(X_1, \ldots, X_n)).$$

DEFINITION 4.2. $S(x_1, \ldots x_n; R)$ *is an* unbiased confidence region *for* $\xi_p(F)$ *if*

$$\text{(4.7)} \qquad P_F(\theta) \geqslant P_F(\xi_p(F))$$

for all distributions $F \in \mathfrak{F}_0$.

The following theorems give us confidence regions for $\xi_p(F)$ with optimum properties.

THEOREM 4.3. *A most powerful (one-sided) similar confidence region for* $\xi_p(F)$ *is* $]-\infty, u(x_1, \ldots, x_n; R)[$ *where*

$$(4.8)\qquad u(x_1, \ldots, x_n; R) = x_{(i)} \quad \textit{with probability } 1-\alpha$$
$$= x_{(i+1)} \quad \textit{with probability } \alpha.$$

This confidence region has maximum power for $\theta > \xi_p(F)$ *among all similar confidence regions.*

Proof. Similar confidence regions must have $P_F(\xi_p(F)) = 1 - \beta$. For $\theta = 0$ and $\xi_p(F) = 0$, we obtain

$$E_F(\phi_0(X_1, \ldots, X_n)) \equiv \beta.$$

Lemma (2.1) gives the restriction

$$(4.9)\qquad \sum_{x=z,y} q^{i(x_1,\ldots,x_n)} p^{n-i(x_1,\ldots,x_n)} \phi_0(x_1, \ldots, x_n) = \beta \qquad \text{a.e.}$$

on $\phi_0(x_1, \ldots, x_n)$. For a distribution $F(x)$ having $F(0) = P_F = 1 - Q_F$, let the probability density $f(x) = P_F f_-(x) + Q_F f_+(x)$ where $f_+(x)$ and $f_-(x)$ are as defined in Lemma 2.1. Setting

$$F_+(x) = \int_0^x f_+(x)\, dx, \quad F_-(x) = \int_{-\infty}^x f_-(x)\, dx,$$

the power function of the confidence region satisfies

$$(4.10)\quad 1 - P_F(0) = \int_{[0,1]^n} \sum_{x=G_-^{-1}, G_+^{-1}} Q_F^{\,i(x(u_1),\ldots,x(u_n))} P_F^{\,n-i(x(u_1),\ldots,x(u_n))} \phi_0(x(u_1), \ldots, x(u_n)) \prod_1^n du_i.$$

Since we are considering the value of $\theta = 0$, our problem is to maximize the power for $0 \geqslant \xi_p(F)$, that is, for $p \leqslant P_F$. A solution to this maximization is obtained by minimizing the integrand of (4.10) subject to (4.9) and the restriction that the values of ϕ_0 belong to [0,1]. This is a simple binomial distribution problem with solution

$$\phi_0(x_1, \ldots, x_n) = 0 \qquad \text{if } i(x_1, \ldots, x_n) < n - i,$$
$$= \alpha \qquad \text{if } i(x_1, \ldots, x_n) = n - i,$$
$$= 1 \qquad \text{if } i(x_1, \ldots, x_n) > n - i.$$

for some α, i.

It will be derived as a corollary to Theorem (5.1) *that the confidence region* (4.8) *is a most powerful* (*one-sided*) β *confidence region.*

THEOREM 4.4. *A most powerful unbiased confidence region for* $\xi_p(F)$ *is*

$$(4.11)\quad [f(x_1, \ldots, x_n; R), g(x_1, \ldots, x_n; R)] = [x_{(i)}, x_{(i+j)}] \quad \textit{with probability } p_1,$$
$$= [x_{(i+1)}, x_{(i+j)}] \quad \textit{with probability } p_2,$$
$$= [x_{(i)}, x_{(i+j+1)}] \quad \textit{with probability } p_3,$$
$$= [x_{(i+1)}, x_{(i+j+1)}] \quad \textit{with probability } p_4,$$

where i, j, p_1, p_2, p_3, p_4 are chosen to make the interval unbiased with confidence level β. This confidence region has maximum power for $\theta \neq \xi_p(F)$ among unbiased similar confidence regions.

Proof. The proof using Lemma (2.1) follows closely that used in Theorem 4.3. It is worth noting that there remains one degree of freedom in the choice of the p's.

That the confidence region (4.11) *is a most powerful unbiased β confidence region for $\xi_p(F)$ will be derived as a corollary to Theorem* (5.2).

5. Tests for the location parameter. We obtain most powerful and most powerful unbiased tests for some hypotheses concerning the location parameter. Consider first a hypothesis completely specifying the location parameter.

$$\text{Hypothesis 1: } \xi_p(F) = \theta,\ F \in \mathfrak{F}_0;$$
$$\text{Alternative 1: } \xi_p(F) > \theta,\ F \in \mathfrak{F}_0.$$

For the problem of obtaining a test, we have the following

THEOREM 5.1. *The one-sided sign test applied to $(x_1 - \theta, \ldots, x_n - \theta)$ is most powerful for the* Hypothesis 1 *against the* Alternative 1.

Proof. For simplicity assume $\theta = 0$ and consider a distribution belonging to the Alternative 1. It will have a density function $f(x)$ which can be decomposed into $f_-(x)$ and $f_+(x)$ as in Lemma (2.1) giving

$$f(x) = p'f_-(x) + q'f_+(x).$$

Following a procedure used by Lehmann in **(4)**, we look for a distribution over the parameter space of Hypothesis 1. The obvious choice for such a least favourable distribution is to give probability one to the distribution $f_0(x) = pf_-(x) + qf_+(x)$. For a sample of n the most powerful test of $f_0(x)$ against $f(x)$ is given by the test function.

$$\phi(x_1, \ldots, x_n) = \begin{cases} 1 & \text{if } \dfrac{\prod f(x_i)}{\prod f_0(x_i)} > c, \\ \alpha & \text{if } \dfrac{\prod f(x_i)}{\prod f_0(x_i)} = c, \\ 0 & \text{if } \dfrac{\prod f(x_i)}{\prod f_0(x_i)} < c. \end{cases}$$

Obviously this is equivalent to

$$\phi(x_1, \ldots, x_n) = \begin{cases} 1 & \text{if } i(x_1, \ldots, x_n) > i_0, \\ \alpha & \text{if } i(x_1, \ldots, x_n) = i_0, \\ 0 & \text{if } i(x_1, \ldots, x_n) < i_0. \end{cases}$$

More generally the sign test is based on the $(x_i - \theta)$'s.

COROLLARY. *The one-sided sign test is most powerful for the* Hypothesis 2 against the *Alternative* 2.

$$\text{Hypothesis 2: } \xi_p(F) \in S,\ F \in \mathfrak{F}_0$$
$$\text{Alternative 2: } \xi_p(F) \in S',\ F \in \mathfrak{F}_0.$$

S and S' are sets on R^1 having $\sup S \leqslant \inf S'$. The test is based on the signs of $x_1 - \sup S, \ldots, x_n - \sup S$.

Proof. Follows easily from Theorem 5.2.

COROLLARY. *The confidence region* (4.8) *for* $\xi_p(F)$ *is a most powerful* (*one-sided*) β *confidence region.*

Proof. By straightforward analogy from the Theorem.

THEOREM 5.2. *The unbiased two-sided sign test is most powerful unbiased for*

$$\text{Hypothesis 3: } \xi_p(F) = 0$$
$$\text{Alternative 3: } \xi_p(F) \neq 0.$$

For the proof of this theorem the following lemma is needed.

LEMMA 5.1. *Any unbiased test of* Hypothesis 3 *against* Alternative 3 *is a test similar over* Hypothesis 3.

Proof. Consider a distribution belonging to Hypothesis 3 and having a continuous density function $f(x)$ for which $f(0) > 0$ and

$$f(x + \epsilon) \leqslant G(x)\,(|\epsilon| < \delta)$$

where $G(x)$ is integrable. The power of an unbiased test $\phi(x_1, \ldots, x_n)$ is

$$\int \phi(x_1, \ldots, x_n) \prod f(x_i + \epsilon) \prod dx_i,$$

and is a continuous function of ϵ. Since we have assumed $\phi(x_1, \ldots, x_n)$ to be unbiased of, say, size α, we have

$$\int \phi(x_1, \ldots, x_n) \prod f(x_i + \epsilon) \prod dx_i \geqslant \alpha \qquad \text{if } \epsilon \neq 0.$$

From the continuity we obtain

$$\int \phi(x_1, \ldots, x_n) \prod f(x_i) \prod dx_i \geqslant \alpha;$$

but since $f(x)$ corresponds to a distribution of Hypothesis 3 we have that the above expression is less than or equal to α. Therefore

$$\int \phi(x_1, \ldots, x_n) \prod f(x_i) \prod dx_i = \alpha$$

for all distributions belonging to Hypothesis 3 which have a continuous density satisfying the bounding condition. Such a class of distributions can replace $\mathfrak{F}_0(p)$ in Theorem 2.1 with the results remaining valid. Hence we have conditions on $\phi(x_1, \ldots, x_n)$ and obtain that it is a test similar over Hypothesis 3. This proves the lemma. This type of argument from unbiasedness to similarity was used by Lehmann.

Proof of Theorem 5.2. From Lemma 5.1 any unbiased test is a similar test and hence by Lemma 2.1 has the form (2.4). Following the argument used in Theorem 4.4, we obtain a solution test function

$$\begin{aligned}\phi(x_1, \ldots, x_n) &= 1 \quad \text{if } x_{(i+j+1)} < 0, \\ &= c' \quad \text{if } x_{(i+j)} < 0 < x_{(i+j+1)}, \\ &= 0 \quad \text{if } x_{(i+1)} < 0 < x_{(i+j)}, \\ &= c'' \quad \text{if } x_{(i)} < 0 < x_{(i+1)}, \\ &= 1 \quad \text{if } 0 < x_{(i)},\end{aligned}$$

where i, j, c', c'' are chosen to make the test unbiased of size α. This completes the proof.

COROLLARY. *The confidence region* (4.11) *is a most powerful unbiased* β *confidence region for* $\xi_p(F)$.

Proof. By straightforward analogy from the Theorem.

THEOREM 5.3. *The most powerful unbiased test of* Hypothesis 3 *against* Alternative 3 *is most stringent if* $p = \frac{1}{2}$.

A test is most stringent if it minimizes the maximum difference between envelope power and power. Thus if ϕ (or ϕ') is a test function of size α for a hypothesis H, then ϕ^* also of size α for H is most stringent against the alternative hypothesis $\bar{H}$ if ϕ^* minimizes

$$\sup_{F \in \bar{H}} \left(\sup_{\phi'} E_F(\phi') - E_F(\phi) \right)$$

as a function of ϕ.

Proof. Let $f_1(x)$ be the density of any distribution belonging to Alternative 3. Then $f_1(x) = p'f_-(x) + q'f_+(x)$. $f_2(x) = q'f_-(x) + p'f_+(x)$ is also the density of a distribution belonging to Alternative 3. From symmetry and Theorem 5.1, we know that the envelope of the power functions for size α tests of Hypothesis 3 has the same value for these two distributions.

For a least favourable distribution over Hypothesis 3 we would choose all probability for the distribution having density $f(x) = \frac{1}{2}f_-(x) + \frac{1}{2}f_+(x)$; and for a distribution over the two alternatives mentioned above we would take probabality $\frac{1}{2}$ for each. A most powerful test for this reduced problem is

$$\begin{aligned}\phi(x_1, \ldots, x_n) &= 1 \quad \text{if } \frac{\frac{1}{2}\prod f_1(x_i) + \frac{1}{2}\prod f_2(x_i)}{\prod f(x_i)} > K, \\ &= c \quad \text{if } \frac{\frac{1}{2}\prod f_1(x_i) + \frac{1}{2}\prod f_2(x_i)}{\prod f(x_i)} = K, \\ &= 0 \quad \text{if } \frac{\frac{1}{2}\prod f_1(x_i) + \frac{1}{2}\prod f_2(x_i)}{\prod f(x_i)} < K.\end{aligned}$$

But

$$\frac{\frac{1}{2}\prod f_1(x_i) + \frac{1}{2}\prod f_2(x_i)}{\prod f(x_i)} = \frac{1}{2}\left(\frac{p'}{\frac{1}{2}}\right)^{n-i}\left(\frac{q'}{\frac{1}{2}}\right)^{i} + \frac{1}{2}\left(\frac{q'}{\frac{1}{2}}\right)^{n-i}\left(\frac{p'}{\frac{1}{2}}\right)^{i},$$

where $i = i(x_1, \ldots, x_n)$. From this it is easily seen that the test is the one obtained in the previous theorem.

This test is similar, it maximizes the minimum power for the two simple alternatives (since the power is the same for these two alternatives), and the test is independent of the alternatives used in its derivation. By the theorem of Hunt and Stein **(5)**, the test is most stringent.

6. A bivariate problem. A familiar statistical problem is the following. Observations are obtained in pairs (x_i, y_i). The x_i and y_i values come from the same, say plot, and the y value is the result corresponding to some "treatment" while the x value corresponds to no treatment. The problem is to find whether the y values tend to be larger than the x values. Often the x and y components cannot be assumed independent, and perhaps no assumption can be made concerning the joint distribution.

In such a situation one or other of the following formulations might be a suitable idealization of the problem. $F(x,y)$ has a density $f(x,y)$:

Hypothesis I: $\xi_{.5}(F(\infty,y)) - \xi_{.5}(F(x,\infty)) = 0$,
Alternative I: $\xi_{.5}(F(\infty,y)) - \xi_{.5}(F(x,\infty)) > 0$.
Hypothesis II: $\xi_{.5}(G(z)) = 0$,
Alternative II: $\xi_{.5}(G(z)) > 0$,

where $G(z)$ is the distribution function corresponding to $y - x$.

It seems to the author that second formulation is more realistic. In any case it is the second formulation for which this paper gives an answer.

THEOREM 6.1. *For a sample of n from a distribution $F(x,y)$ having a density, the one-sided sign test is most powerful for formulation* II.

Proof. Consider an alternative distribution having density $f_1(x,y)$ and let

$$f_1(x, y) = pf_-(x, y) + qf_+(x, y),$$

where $f_-(x,y)$ and $f_+(x,y)$ are respectively density functions over the regions $\{(x,y)|y - x < 0\}$, $\{(x,y)|y - x > 0\}$. Obviously $p < \frac{1}{2}$.

The distribution with density $f_0(x,y) = \frac{1}{2}f_-(x,y) + \frac{1}{2}f_+(x,y)$ belongs to the Hypothesis. Giving this distribution probability one as a least favourable distribution over the Hypothesis, we obtain the test

$$\begin{aligned}\phi(x_i, y_i) &= 1 &&\text{if } \frac{\prod f_1(x_i, y_i)}{\prod f_0(x_i, y_i)} > c,\\ &= k &&\text{if } \frac{\prod f_1(x_i, y_i)}{\prod f_0(x_i, y_i)} = c,\\ &= 0 &&\text{if } \frac{\prod f_1(x_i, y_i)}{\prod f_0(x_i, y_i)} < c.\end{aligned}$$

https://doi.org/10.4153/CJM-1954-007-7 Published online by Cambridge University Press

This is the one-sided sign test mentioned in the statement of the theorem. Since it is a similar test, it is a most powerful test of the Hypothesis against the particular alternative. However the test does not depend on the alternative used in the derivation; hence it is a uniformly most powerful test for formulation II.

The sign test utilizes only the signs of the differences $y_i - x_i$. A test based on the signs of the differences $y_i - x_i$ and on the ranks of the numbers $|y_1 - x_1|, \ldots, |y_n - x_n|$ was proposed by Wilcoxen **(9)**; the procedure for applying the test is similar to that for the Wilcoxen (Mann-Whitney) two sample test. This sign-rank test is designed to test the more restricted hypothesis:

Hypothesis III: $f(x,y)$ is symmetric about $y - x = 0$. This hypothesis requires that $\xi_{.5}(G(z)) = 0$ and in addition that the distribution is symmetric about $z = 0$.

For the restricted Hypothesis III, conceivably the Wilcoxen sign-rank test (of size α) could be more powerful for certain alternatives than the sign test (of size α).

For the Hypothesis II the Wilcoxen sign-rank test does not apply (the size determination in **(9)** presupposes Hypothesis III), and the sign test as was shown above is most powerful.

7. Definition of scale and location-scale parameters. Conditions that parameters be scale and location-scale parameters for a class of distributions as large as $\mathfrak{F}_0$ could be formulated along the lines followed for the location parameters in §3. The result however would be the definitions:

DEFINITION (7.1). *The* scale parameter *is* $\eta(F) = \xi_{p_2}(F) - \xi_{p_1}(F)$.

DEFINITION (7.2). *The* location-scale parameter *is* $(\xi_{p_1}(F), \xi_{p_2}(F))$.

8. Confidence regions and tests for the scale parameter. It is not difficult to find confidence intervals for the scale parameter; they may be derived from the order statistics. No attempt is made to get a best confidence interval, but rather a result of negative nature is obtained; the nonexistence of similar confidence regions.

THEOREM 8.1. *Similar β ($\beta \neq 0,1$) confidence regions do not exist for $\eta(F)$ (other than degenerate regions).*

To prove this theorem we need an analogue of Theorem 2.1 for distributions having two percentiles fixed.

THEOREM 8.2. *For the class of distributions*

$$\mathfrak{F}_{ab}^n = \{\prod F(x_i) | F(x) \in \mathfrak{F}_0, \xi_{p_1}(F) = a, \xi_{p_2}(F) = b\}$$

a symmetric unbiased estimate of zero has the form

$$(8.1) \qquad \phi(x_1, \ldots, x_n) = \sum \alpha(x_i)\, \psi_a(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) + \sum \beta(x_i)\, \psi_b(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n),$$

https://doi.org/10.4153/CJM-1954-007-7 Published online by Cambridge University Press

where

$$\begin{aligned} \beta(x) &= +p_2 && \text{if } x > b, \\ &= -(1-p_2) && \text{if } x < b, \\ \alpha(x) &= +p_1 && \text{if } x > a, \\ &= -(1-p_1) && \text{if } x < a, \end{aligned}$$

and ψ_a and ψ_b are bounded and symmetric.

Proof. Although the proof of this theorem can be given quite similarly to that of Theorem 2.1, we outline another form, the steps of which can be used in the proof of Theorem 8.1.

If $f_1(x), \ldots, f_n(x)$ are arbitrary bounded measurable functions, we define $\phi(f_1, \ldots, f_n)$ by

$$\phi(f_1, \ldots, f_n) = \int \phi(x_1, \ldots, x_n) \prod f_i(x_i) \prod dx_i.$$

Since $\phi(f, \ldots, f) = 0$ if $f \in \mathfrak{F}_{ab}$, it follows by the method of proof in **(3)** that $\phi(f_1, \ldots, f_n) = 0$ if all $f_i \in \mathfrak{F}_{ab}$.

Defining $\alpha(f)$, $\beta(f)$ in the manner used for $\phi(f_1, \ldots, f_n)$, we have $\phi(f_1, \ldots, f_n) = 0$ if $\alpha(f_1) = \beta(f_1) = \ldots = \beta(f_n) = 0$ where the f's may be linear combinations of elements of $\mathfrak{F}_0$. If f_0^α and f_0^β have $\alpha(f_0^\alpha) \neq 0$, $\beta(f_0^\alpha) = 0$ and $\alpha(f_0^\beta) = 0$, $\beta(f_0^\beta) \neq 0$, then

$$\phi(f, f_2, \ldots, f_n) - \frac{\alpha(f)}{\alpha(f_0^\alpha)} \phi(f_0^\alpha, f_2, \ldots, f_n) - \frac{\beta(f)}{\beta(f_0^\beta)} \phi(f_0^\beta, f_2, \ldots, f_n)$$

will be zero if $\alpha(f_2) = \beta(f_2) = \ldots = \beta(f_n) = 0$. This obtains from a simple analysis of linear functions over a vector space. Proceeding in this manner, the expression (8.1) is obtained fairly easily.

Proof of Theorem 8.1. Letting $S(x_1, \ldots, x_n; R)$ be a confidence region for $\eta(F)$, we define a corresponding characteristic function.

$$\phi_\eta(x_1, \ldots, x_n) = \frac{1}{n!} \sum \Pr_R\{\eta \in S(x_{i_1}, \ldots, x_{i_n}; R)\},$$

where the summation is over the $n!$ permutations. If S is a similar confidence region, then

$$E_F\{\phi_\eta(X_1, \ldots, X_n)\} = \beta$$

for all distributions belonging to $\mathfrak{F}_{x_0 x_0 + \eta}$. Thus

$$\phi_\eta(x_1, \ldots, x_n) - \beta = \overset{*}{\phi_\eta}(x_1, \ldots, x_n)$$

is an unbiased estimate of zero for $F \in \mathfrak{F}_{x_0 x_0 + \eta}$ for all x_0. Letting $\alpha_{x_0}(x)$, $\beta_{x_0}(x)$ be as defined in Theorem 8.2 with $a = x_0$ and $b = x_0 + \eta$, then

$$\phi^*(f, f_2, \ldots, f_n) - \frac{\alpha_{x_0}(f)}{\alpha_{x_0}(f_0^\alpha)} \phi^*(f_0^\alpha, f_2, \ldots, f_n) - \frac{\beta_{x_0}(f)}{\beta_{x_0}(f_0^\beta)} \phi^*(f_0^\beta, f_2, \ldots, f_n)$$

is equal to zero whenever

https://doi.org/10.4153/CJM-1954-007-7 Published online by Cambridge University Press

$$\alpha_{x_0}(f_2) = \beta_{x_0}(f_2) = \ldots = \beta_{x_0}(f_n) = 0.$$

If in addition $f_2, \ldots, f_n, f_0^\alpha, f_0^\beta$ are functions equal to zero in the intervals $[a_0, a_0 + \epsilon]$, $[a_0 + \eta, a_0 + \eta + \epsilon]$, then by taking f to be nonzero only in $[a_0, a_0 + \frac{1}{2}\epsilon]$ or in $[a_0 + \eta + \frac{1}{2}\epsilon, a_0 + \eta + \epsilon]$ and by changing x_0 from a_0 to $a_0 + \frac{1}{2}\epsilon$ to $a_0 + \epsilon$, the second and third terms are changed successively in sign without altering the value of the expression as a whole. Hence $\phi^*(f_0^\alpha, f_2, \ldots, f_n)$ and $\phi^*(f_0^\beta, f_2, \ldots, f_n)$ are equal to zero if $\alpha_a(f_2) = \beta_a(f_2) = \ldots = \beta_a(f_n) = 0$ and if these functions are zero in $(a, a + \epsilon]$ and $[a + \eta, a + \eta + \epsilon]$.

Proceeding in this manner we obtain finally that if $f_0^{(1)}, \ldots, f_0^{(n)}$ are equal to 1 respectively on the disjoint intervals $I_1, \ldots, I_n$ and are zero elsewhere, then

$$\int_{\prod I_i} \phi^*(x_1, \ldots, x_n) \prod dx_i = 0$$

and $\phi^*(x_1, \ldots, x_n) = 0$ almost everywhere.

Thus we obtain $\phi^\eta \equiv \beta$. This means that given the order statistics $x_{(1)}, \ldots, x_{(n)}$ the probability is β that $S(x_1, \ldots, x_n; R)$ covers any arbitrary positive real number. Such a confidence region is essentially equivalent to

$$S(x_1, \ldots, x_n; R) =]0, \infty[\qquad \text{if } R \leqslant \beta,$$
$$= \phi \qquad \text{if } R > \beta,$$

which we refer to as a degenerate confidence interval. This completes the proof.

By analogy we have

THEOREM 8.3. *For the hypothesis* $\eta(F) = \eta_0$, *similar tests other than* $\phi(x_1, \ldots, x_n) \equiv \alpha$ *do not exist.*

9. Distribution-free upper tolerance bounds. A distribution-free upper tolerance bound is a particular type of distribution-free or non-parametric tolerance region. We first define this latter concept. Let $\mathscr{S}$, $\mathfrak{A}$ be a measurable space, that is, $\mathscr{S}$ is an arbitrary space and $\mathfrak{A}$ is a class of subsets A of $\mathscr{S}$ which form a σ-algebra, and let $\{P_\theta(A) | \theta \in \Omega\}$ be a class of probability measures over the space $\mathscr{S}$.

A tolerance region for the class of measures $\{P_\theta(A) | \theta \in \Omega\}$ *is a function* $A(x_1, \ldots, x_n)$ *which maps* $\mathscr{S}^n$ *into the class* $\mathfrak{A}$ *and for which the distribution of* $P_\theta(A(x_1, \ldots, x_n))$ *induced by the product probability measure* P_θ^n *over* $\mathscr{S}^n$ *is independent of* $\theta \in \Omega$.

Weaker forms of this definition have been used in particular problems. For a distribution-free upper tolerance bound we need $\mathscr{S} = R^1$, $\{P_\theta(A) | \theta \in \Omega\} = \mathfrak{F}_0$, and $A(x_1, \ldots, x_n) =]-\infty, f(x_1, \ldots, x_n)[$.

In 1944 Robbins **(6)** considered the problem of finding the most general distribution-free upper tolerance bounds. He proved that, subject to continuity restrictions on the function $f(x_1, \ldots, x_n)$,

$$\prod_1^n [f(x_1, \ldots, x_n) - x_i] = 0.$$

https://doi.org/10.4153/CJM-1954-007-7 Published online by Cambridge University Press

Here we remove the continuity restrictions and envisage randomized bounds $f(x_1, \ldots, x_n; R)$ where R is a random variable with a uniform distribution on [0,1]. Our result is in effect that $f(x_1, \ldots, x_n; R)$ chooses the order statistics with fixed probabilities.

The problem of the most general bound can be given quite interestingly as a measure problem. In **(7)** it was shown that any continuous distribution over R^1 can be obtained from the uniform distribution over [0,1] by a monotone strictly increasing mapping of [0,1] into R^1, and conversely. Let such a mapping be $g(u)$, corresponding to the distribution function $G(x)$. Then essentially $g(u) = G^{-1}(u)$ where G^{-1} is the inverse function of $G(x)$. Also let $\mathfrak{G}$ be the class of all continuous distribution functions $G(x)$. Then to find the distribution-free upper tolerance bounds is to find functions $f(x_1, \ldots, x_n; R)$ for which the Lebesgue measure of

$$\{(u_1, \ldots, u_n, u_{n+1}) | G(f(G^{-1}(u_1), \ldots, G^{-1}(u_n); u_{n+1})) \leqslant v, \quad (u_1, \ldots, u_{n+1}) \in (0,1]^{n+1}\}$$

is independent of $G(x) \in \mathfrak{G}$, for all $v \in [0,1]$.

THEOREM 9.1. *A necessary and sufficient condition that* $f(x_1, \ldots, x_n; R)$ be a *distribution-free upper tolerance limit is that* $f(x_{i_1}, \ldots, x_{i_n}; R)$ *chosen with probability* $1/n!$ *for fixed* $x_{(1)}, \ldots, x_{(n)}$ *should be equivalent to* $x_{(1)}, \ldots, x_{(n)}$, *chosen with fixed but arbitrary probabilities* $p_1, \ldots, p_n$ *respectively where* $\sum p_i = 1$ (*almost everywhere* $\mathfrak{G}$).

Proof. Let $H(v)$ be the distribution function of

$$P_G(]-\infty, f(X_1, \ldots, X_n; R)[) = G(f(X_1, \ldots, X_n; R))$$

when the $X_1, \ldots, X_n$ have the distribution function $G(x) \in \mathfrak{G}$. Then

$$\begin{aligned} H(v) &= \Pr_G\{G(f(X_1, \ldots, X_n; R)) \leqslant v\} \\ &= \Pr_G\{f(X_1, \ldots, X_n; R) \leqslant G^{-1}(v)\} \\ &= \Pr_G\{f(X_1, \ldots, X_n; R) \leqslant \xi_v(G)\}. \end{aligned}$$

Thus $f(x_1, \ldots, x_n; R)$ is a similar $1 - H(v)$ confidence bound for $\xi_v(G)$. Letting $\phi_\theta(x_1, \ldots, x_n)$ be the characteristic function of the region $]-\infty, f(x_1, \ldots, x_n; R)[$, Lemma 2.1 gives conditions on ϕ_θ which for $\theta = 0$ are

$$\sum_{x=z,y} v^{n-i(x_1,\ldots,x_n)}(1-v)^{i(x_1,\ldots,x_n)}\phi_0(x_1, \ldots, x_n) = 1 - H(v)$$

where the $z_i < 0$ and the $y_i > 0$. Not only must this hold almost everywhere $(z_1, \ldots, z_n; y_1, \ldots, y_n) \in]-\infty, 0[^n \times]0, \infty[^n$, but it must hold for all v. The above equation determines the form of the right hand side

$$1 - H(v) = \sum C_i v^{n-i}(1-v)^i,$$

and this in turn implies

$$\phi_0(x_1, \ldots, x_n) = c_{i(x_1,\ldots,x_n)}$$

https://doi.org/10.4153/CJM-1954-007-7 Published online by Cambridge University Press

almost everywhere. Similarly we obtain

$$\phi_\theta(x_1, \ldots, x_n) = c_{i(x_1-\theta, \ldots, x_n-\theta)}.$$

Setting

$$c_i = \sum_{j=0}^{i} p_j$$

and assuming $f(x_1, \ldots, x_n; R)$ to be symmetric in the x's, we obtain

$$\begin{aligned} f(x_1, \ldots, x_n; R) &= +\infty \quad \text{with probability } p_0, \\ &= x_{(n)} \quad \text{with probability } p_1, \\ &= x_{(1)} \quad \text{with probability } p_n, \\ &= -\infty \quad \text{with probability } 1 - \sum_0^n p_i. \end{aligned}$$

Then with the obvious modification if $f(x_1, \ldots, x_n; R)$ is not symmetric, the theorem is proved.

10. Two theorems in estimation theory. In their 1950 paper **(8)** Lehmann and Scheffé give several theorems for unbiased estimation. One of these defines the class of uniformly minimum variance (UMV) unbiased estimates given the class of unbiased estimates of zero. For non-parametric application this was restricted in that they considered only estimates with finite variance over the parameter space. Since there are reasonable non-parametric estimates not satisfying this condition, consider an extension of their results.

Let $\{P_\theta^x | \theta \in \Omega\}$ be the class of distributions under consideration and let $T(x)$ be a sufficient statistic with corresponding distributions $\{P_\theta^T | \theta \in \Omega\}$. We now consider the estimation of real valued functions $g^*(\theta)$ which exist over the parameter space.

LEMMA 10.1. *Under the assumption that all statistics in ν_0 have finite variance, a statistic is a minimum variance* (UMV) *unbiased estimate of its expected value if and only if it belongs to ν_1 where*

$$\nu_0 = \{f(t) | E_\theta\{f(T)\} \equiv 0, \theta \in \Omega\},$$

$$\nu_1 = \{g(t) | E_\theta\{g(T)f(T)\} = 0 (\theta \in \Omega, f(t) \in \nu_0, \text{Var}_\theta g(T) < \infty); \quad E_\theta\{g(T)\} < \infty \ (\theta \in \Omega)\}.$$

Proof. Since we are concerned with UMV unbiased estimates, the Rao-Blackwell theorem says that we may restrict attention to estimates based on the sufficient statistic.

Let $g(t)$ be a UMV unbiased estimate of $g^*(\theta)$. If $f(t) \in \nu_0$ then $g(t) + \lambda f(t)$ is also unbiased for $g^*(\theta)$ and must have variance at least as large as $g(t)$.

$$\begin{aligned} \text{Var}_\theta\{g(T) + \lambda f(T)\} &= \text{Var}_\theta\{g(T)\} + 2\lambda E_\theta\{g(T)f(T)\} + \lambda^2 \text{Var}_\theta\{f(T)\} \\ &\geqslant \text{Var}_\theta\{g(T)\}. \end{aligned}$$

https://doi.org/10.4153/CJM-1954-007-7 Published online by Cambridge University Press

If $\mathrm{Var}_\theta\{g(T)\}$ is finite then the above inequality being true for all positive and negative λ implies that $E_\theta\{g(T)f(T)\} = 0$, that is $g(t) \in \nu_1$.

Next assume that $g(t) \in \nu_1$ and let $g'(t)$ be any other unbiased estimate of $g^*(\theta)$. Then $g'(t) - g(t)$ is an unbiased estimate of zero, say $f(t)$, and

$$\begin{aligned}\mathrm{Var}_\theta\{g'(T)\} &= \mathrm{Var}_\theta\{g(T) + f(T)\} \\ &= \mathrm{Var}_\theta\{g(T)\} \qquad\qquad\qquad \text{if } \mathrm{Var}_\theta\{g(T)\} = \infty, \\ &= \mathrm{Var}_\theta\{g(T)\} + \mathrm{Var}_\theta\{f(T)\} \text{ if } \mathrm{Var}_\theta\{g(T)\} < \infty.\end{aligned}$$

In either case we have $\mathrm{Var}_\theta\{g'(T)\} \geqslant \mathrm{Var}_\theta\{g(T)\}$ which means that $g(t)$ is a UMV unbiased estimate of $g^*(\theta)$. This proves the lemma.

Also for convex loss functions we have the following

LEMMA 10.2. *If a real valued parameter $g(\theta)$ has a minimum risk unbiased estimate, then the estimate is unique almost everywhere $\{P_\theta^x\}$ (assuming the loss function is strictly convex and the risk finite).*

Proof. Let $W_\theta(f)$ be the strictly convex loss function; then

$$\alpha W_\theta(f) + (1-\alpha)\, W_\theta(f') \geqslant W_\theta(\alpha f + (1-\alpha)f')$$

if $\alpha \in\,]0,1[$ and $f \neq f'$. Suppose $f(t)$ and $f'(t)$ are minimum risk unbiased estimates of $g(\theta)$ with

$$h(\theta) = E_\theta(W_\theta(f)) = E_\theta(W_\theta(f')).$$

It follows that $\alpha f + (1-\alpha)f'$ is an unbiased estimate of $g(\theta)$. Since f and f' are minimum risk estimates, then for $\alpha \in\,]0,1[$

$$E_\theta\{W_\theta(\alpha f + (1-\alpha)f')\} \geqslant h(\theta);$$

but

$$E_\theta\{W_\theta(\alpha f + (1-\alpha)f')\} \leqslant E_\theta\{\alpha W_\theta(f) + (1-\alpha)\, W_\theta(f')\}$$

with inequality strict unless

$$W_\theta(\alpha f + (1-\alpha)f') = \alpha W_\theta(f) + (1-\alpha)\, W_\theta(f')$$

almost everywhere which implies $f = f'$ almost everywhere. But by combining the inequalities we see that they are equalities and hence $f = f'$ almost everywhere. This proves the lemma.

11. Some examples of estimation in non-parametric theory. For a sample $X_1, \ldots, X_n$ from an unknown distribution assumed to be absolutely continuous, the order statistics $x_{(1)}, \ldots, x_{(n)}$ form a complete sufficient statistic. However, the Lehmann-Scheffé theorem on minimum variance and minimum risk unbiased estimation can not be applied immediately in most cases.

Consider the estimation of $E_\theta(X^2)$. The essential step in the Lehmann-Scheffé theorem is in showing that the estimate which depends on $x_{(1)}, \ldots x_{(n)}$ is unique. Let $f_1(x_{(1)}, \ldots x_{(n)}$ and $f_2(x_{(1)}, \ldots, x_{(n)})$ be two such unbiased estimates; then $f_1 - f_2$ is an unbiased estimate of zero for those distributions for which $E_\theta(X^2)$

https://doi.org/10.4153/CJM-1954-007-7 Published online by Cambridge University Press

is finite. Both $\mathfrak{F}_2$ and $\mathfrak{F}_3$ consist of such distributions; hence $f_1 = f_2$ almost everywhere. Thus it is essential to check that the sufficient statistic is complete for the distributions for which the parameter in question exists.

As a second example we consider the problem of obtaining minimum variance unbiased estimates of the parameters of the distributions $\mathfrak{F}_0(p)$ $(p \neq 0, 1)$. We restrict attention to parameters $g(\theta)$ which exist at least for a minimal class sufficiently large that the class of unbiased estimates of zero contains only estimates with finite variance and we apply Lemma 10.1. ν_0 contains statistics $f(x_{(1)}, \ldots, x_{(n)})$ satisfying (2.4).

A statistic $g(x_{(1)}, \ldots, x_{(n)})$ in ν_1 will satisfy

$$E\{g(x_{(1)}, \ldots, x_{(n)}) f(x_{(1)}, \ldots, x_{(n)})\} = 0$$

whenever Var $g < \infty$. For those statistics having finite variance for the minimal class of distributions mentioned above, $f(x_{(1)}, \ldots, x_{(n)})\ g(x_{(1)}, \ldots, x_{(n)})$ will also satisfy (2.4). Thus for every $f(x_{(1)}, \ldots, x_{(n)})$ satisfying

$$\sum_{x=z,y} \left(\frac{q}{p}\right)^{i(x_1, \ldots, x_n)} f(x_1, \ldots, x_n) = 0,$$

we have

$$\sum_{x=z,y} \left(\frac{q}{p}\right)^{i(x_1, \ldots, x_n)} f(x_1, \ldots, x_n)\, g(x_1, \ldots, x_n) = 0.$$

It follows that $g(x_1, \ldots, x_n) = 0$ and hence that there are no nondegenerate minimum variance unbiased estimates which have finite variance over the minimal class of distributions.

References

1. P. Halmos, *The theory of unbiased estimation*, Ann. Math. Stat., *17* (1946), 34–43.
2. E. L. Lehmann, *Notes on the theory of estimation*, Lecture notes mimeographed at the University of California.
3. D. A. S. Fraser, *Completeness of order statistics*, Can. J. Math., *6* (1954), 42-45.
4. E. L. Lehmann and C. Stein, *Most powerful tests of composite hypotheses. I: Normal distributions*, Ann. Math. Stat., *19* (1948), 495–516.
5. E. L. Lehmann and C. Stein, *On the theory of some non-parametric hypotheses*, Ann. Math. Stat. *20* (1949), 28–45.
6. H. Robbins, *On distribution-free tolerance limits in random sampling*, Ann. Math. Stat., *15* (1944), 214–217.
7. H. Scheffé and J. W. Tukey, *Nonparametric estimation. I: Validation of order statistics*, Ann. Math. Stat., *16* (1945), 187–192.
8. E. L. Lehmann and H. Scheffé, *Completeness, similar regions, and unbiased estimation*, Sankhya, *10* (1950), 305–340.
9. F. Wilcoxen, *Individual comparisons by ranking methods*, Biometrics, *1* (1945), 80–83.

University of Toronto