# ON A MEASURE PROBLEM ARISING IN THE THEORY OF NON-PARAMETRIC TESTS

By Henry Scheffé

*Princeton University*

**1. Introduction.** While the contents of this paper have broader statistical implications, they were motivated by the following problem: Given two samples, $(Y_1, Y_2, \cdots, Y_m)$ and $(Z_1, Z_2, \cdots, Z_n)$ from univariate populations with cumulative distribution functions (c.d.f's) $F(x)$ and $G(x)$, respectively, and given furthermore that $F$ and $G$ are members of a certain class $\Omega$ of c.d.f's, to test the hypothesis that $F = G$. We shall refer to this as "the problem of two samples" [8]. It is an example of what Wolfowitz has called problems of the non-parametric case [8].

For the theory of non-parametric problems the following classification of c.d.f's is appropriate: Let $\Omega_0$ *be the class of all univariate c.d.f's*, that is, the class of all monotone non-decreasing functions $F(x)$ for which $F(-\infty) = 0$, $F(+\infty) = 1$, and $F(x) = F(x + 0)$. For every $F \epsilon \Omega_0$ we may conceive of a corresponding random variable $X$ such that $Pr\{X \leq x\} = F(x)$. For some purposes we may desire to rule out the class $\Omega^{(0)}$ of degenerate c.d.f's given by the formula $F(x) = 0$ for $x < x_0$, $F(x) = 1$ for $x \geq x_0$, where $x_0$ is any real number. Let then $\Omega_1$ *be the class of non-degenerate c.d.f's*, $\Omega_1 = \Omega_0 - \Omega^{(0)}$. *Let* $\Omega_2$ *be the class of all continuous* $F(x)$, and let $\Omega_3$ *be the class of all absolutely continuous* $F(x)$, that is, all $F(x)$ for which there exists a probability density function (p.d.f.) $f(x)$ such that

$$F(x) = \int_{-\infty}^{x} f(\xi)\, d\xi. \tag{1}$$

Finally, let $\Omega_4$ *be the class of all* $F(x)$ *which may be expressed in the form* (1) *with* $f(x)$ *continuous.*

Various solutions of non-parametric problems have been given under the restriction that the c.d.f's belong to one of the classes $\Omega_i$. For example, Kolmogoroff [2] has indicated how a confidence belt for an unknown $F$ may be formed with no assumptions on $F$, that is $F \epsilon \Omega_0$. Wald and Wolfowitz earlier[1] gave a more general solution of the same problem [5], and also of the problem of two samples [6], under the restriction that the c.d.f's are members of $\Omega_2$. The latter problem was considered by Dixon [1] for the c.d.f's in $\Omega_3$. Wilks' theory of tolerance intervals [7] assumes $F \epsilon \Omega_4$. The class $\Omega_1$ has been defined above because it is ordinarily the largest class of statistical interest. We note

$$\Omega_0 \supset \Omega_1 \supset \Omega_2 \supset \Omega_3 \supset \Omega_4. \tag{2}$$

[1] See, however, a still earlier paper by Kolmogoroff [11] in which he gave the distribution theory required for his solution.

It is to be understood throughout that the word "*region*" (also the symbol $w$) *always denotes a Borel set in a k-dimensional* ($k > 1$) *sample space* $W$ (Euclidean). *A "null set" will always mean a Borel set of measure zero.*

Returning now to the problem of two samples, let $m + n = k$, $X_i = Y_i$ ($i = 1, 2, \cdots, m$), $X_i = Z_{i-m}$ ($i = m + 1, \cdots, k$). Denote by $E$ the point $(X_1, \cdots, X_k)$. Proceeding along the lines of the usual parametric theory, we may seek a region $w$ (the "critical region") such that $Pr\{E \epsilon w\}$ is the same constant $\alpha$ ("significance level"; $\alpha \neq 0$ or 1) for all $F$ in a particular class $\Omega_i$ if $F = G$. This raises the following question: Define

$$P(w \mid F) = \int_w dF_k(x_1, \cdots, x_k),$$

where

$$F_k(x_1, \cdots, x_k) = \prod_{j=1}^{k} F(x_j).$$

We shall say that a region $w$ *has the property* $\pi_i$ *if for all* $F \epsilon \Omega_i$, $\alpha = P(w \mid F)$ *is independent of* $F$ *and* $0 < \alpha < 1$. The question then is, for a fixed $i$, how can we characterize regions $w$ with the property $\pi_i$? Partial answers to this question are given in the next section.

In the language of measure theory the question is this: Let $\mu$ be any measure on the real line, such that the measure of the whole line is unity, and form the "power" measure $\mu^k$ in Euclidean $k$-space—that is, the product measure obtained by using $\mu$ on each axis. For certain large classes $C_i$ (corresponding to the $\Omega_i$ defined above, $i = 1, 2, 3, 4$) of measures $\mu$, what can we say about the existence and structure of sets of points in the $k$-space which have the property that their "power" measure is the same for all measures $\mu$ in $C_i$?

**2. Theorems.** Our first theorem tells us that if we want regions $w$ with the desired property, we must restrict $F$ to a smaller class than $\Omega_1$.

THEOREM 1: *There is no* $w$ *with the property* $\pi_1$.

To prove the theorem, suppose the contrary. Then there exists a $w$ for which $P(w \mid F) = \alpha$ for all $F \epsilon \Omega_1$ and $\alpha \neq 0$ or 1. Let $L$ be the line $x_1 = x_2 = \cdots = x_k$, and suppose first there is a point $E_0$ of $L$ in $w$. Let $E_0 = (a, a, \cdots, a)$, and let $F_h(x)$ be any $F \epsilon \Omega_1$ such that $Pr\{X = a \mid F_h\} = h$ ($0 < h < 1$). Then

$$\alpha = P(w \mid F_h) \geq P(E_0 \mid F_h) = Pr\{\text{all } X_j = a \mid F_h\}$$
$$= \prod_{j=1}^{k} Pr\{X_j = a \mid F_h\} = h^k.$$

By hypothesis $\alpha$ is independent of $h$. But $h$ may be chosen arbitrarily close to 1. Hence $\alpha = 1$, a contradiction. If no points of $w$ lie on $L$, the above reasoning applies to $w' = W - w$, since $\alpha' = P(w' \mid F) = 1 - \alpha$ is independent of $F \epsilon \Omega_1$, and $w'$ contains an $E_0$ on $L$, therefore $\alpha' = 1$, $\alpha = 0$.

In order to see what kind of structure might yield a $w$ of the desired type, let us for the moment consider the class $\Omega_3$ of c.d.f.'s. Then there exists a p.d.f. over $W$, namely $f(x_1)f(x_2) \cdots f(x_k)$. For any $f(x)$ and any point[2] $E$, this p.d.f. has the same value at all points $E'$ whose coordinates are permutations of the coordinates of $E$. This suggests that suitable regions $w$ can be built up by considering points $E$ for which no two coordinates are equal and putting a fixed fraction of the set $\{E'\}$ in $w$ in such a way that $w$ is a Borel set. Our next theorem justifies this process for the wider class $\Omega_2$.

Let us say that $w$ *has the structure* $S$ *if for every point* $E = (x_1, \cdots, x_k)$ *with no two coordinates equal,* $M$ *points* $(0 < M < k!)$ *of the set* $\{E'\}$*, obtained by permuting the coordinates of* $E$*, are in* $w$ *and the remaining* $k! - M$ *are not.*[3]

THEOREM 2: *A sufficient condition that* $w$ *have the property* $\pi_2$ *is that it have the structure* $S$.

In proving the theorem it will be convenient to separate the $k!$ points of every set $\{E'\}$ by means of regions $u_i$ $(i = 1, \cdots, k!)$, such that each $u_i$ contains one and only one point of $\{E'\}$. Order the $k!$ permutations of the integers $1, 2, \cdots, k$ in any manner so that $(1, 2, \cdots, k)$ is the first. Let $(p_{i1}, \cdots, p_{ik})$ be the $i$th permutation $(i = 1, 2, \cdots, k!)$ and *define* $u_i$ *as the region* $x_{p_{i1}} < x_{p_{i2}} < \cdots < x_{p_{ik}}$. The collection $\{u_i\}$ is disjoint and covers all of $W$ except the *set* $H$ *of points on hyperplanes* $x_i = x_j$ $(i \neq j)$. The transformation $T_i$ : $x_{p_{i1}} \to x_1, \cdots, x_{p_{ik}} \to x_k$ maps $u_i$ onto $u_1$ in such a way that $F_k$ remains invariant.

Suppose now that $w$ satisfies the conditions of the theorem. The removal of $H \cap w$ from $w$ does not[4] affect $P(w \mid F)$ for any $F \in \Omega_2$. Hence

$$P(w \mid F) = \sum_{i=1}^{k!} P(w \cap u_i \mid F) = \sum_{i=1}^{k!} \int_{w \cap u_i} dF_k$$
$$= \sum_{i=1}^{k!} \int_{u_i} c_{w \cap u_i}(E)\, dF_k ,$$

where $c_S(E)$ denotes the characteristic function of a set $S$, that is, $c_S(E) = 1$ if $E \in S$, 0 otherwise. Next map each of the regions $u_i$ onto $u_1$ by means of $T_i$. $F_k$ is invariant, while $c_{w \cap u_i}(E) \to h_i(E)$ such that $\sum_{i=1}^{k!} h_i(E) = M$ for $E \in u_1$. Then

$$P(w \mid F) = \sum_{i=1}^{k!} \int_{u_1} h_i(E)\, dF_k = \int_{u_1} \sum_{i=1}^{k!} h_i(E)\, dF_k = M \int_{u_1} dF_k .$$

---

[2] Previously $E$ denoted a random point $(X_1, \cdots, X_k)$, now it denotes an arbitrary point $(x_1, \cdots, x_k)$ in the sample space $W$. This will cause no confusion.

[3] Regions of structure $S$ may be regarded as the result of applying R. A. Fisher's randomization process [10] in the most general possible way to the problem of two samples. Special cases of regions with structure $S$ have been considered by Feller [9] and Neyman [12], and are implied by all writers [e.g., 6] who have attacked the problem of two samples by the method of ranks.

[4] This may be seen by writing $P(H \mid F)$ in the form of an integral over $W$ of $c_H(E)\, dF_k$, where $c_H(E)$ is the characteristic function of the set $H$, and applying the Fubini theorem [4].

But

$$1 = P(W \mid F) = \sum_{i=1}^{k!} \int_{u_i} dF_k ,$$

and by use of $T_i$ we find

$$\int_{u_i} dF_k = \int_{u_1} dF_k \qquad (i = 1, \cdots, k!).$$

Hence

$$\int_{u_i} dF_k = 1/k!,$$

and

$$P(w \mid F) = M/k!$$

for all $F \epsilon \Omega_2$. Thus $w$ has the property $\pi_2$.

$H$ is an example of a set in the *class* $N_2$ *of regions* $w$ *for which* $P(w \mid F) = 0$ *for all* $F \epsilon \Omega_2$. Since if regions $w_1$ and $w_2$ differ by a set $w \epsilon N_2$, $P(w_1 \mid F) = P(w_2 \mid F)$ for all $F \epsilon \Omega_2$, we have

Corollary 1: *It is sufficient that w have the property $\pi_2$ if it differs from a region with structure S by a region in $N_2$.*

Defining similarly the class $N_3$ as that class of regions $w$ for which $P(w \mid F) = 0$ for all $F \epsilon \Omega_3$, we see that $N_3$ is precisely the class of null sets.

Corollary 2: *A sufficient condition that w have the property $\pi_3$ is that it have the structure S except for a null set.*

The mildest restriction under which the writer has been able to concoct a necessity proof is that the boundary of $w$ be a null set. This class of regions $w$ includes (to the best of his knowledge) all critical regions heretofore used in practice.

Theorem 3: *For a w whose boundary is a null set, a necessary condition that w have the property $\pi_4$ is that it have the structure S except on a null set.*

Suppose then that $w$ has the property $\pi_4$, and its boundary $B$ is a null set. Let $B_i$ be the transform of $B$ under $T_i$. Let the null set $H'$ be the union of $H$ with all $B_i$ and let $w_1 = w - H'$, $w_2 = (W - w) - H'$. Then $w_1$ and $w_2$ are open sets and $P(w_1 \mid F) = P(w \mid F)$ for all $F \epsilon \Omega_4$. Furthermore for any $E$ either all or none of the points of $\{E'\}$ are in $w_1 \cup w_2$. Now consider any $E_0 \epsilon w_1$ and let $M_0$ be the number of points of $\{E_0'\}$ in $w_1$, so that $k! - M_0$ of $\{E_0'\}$ are in $w_2$. Let $E_0 = (\xi_1, \cdots, \xi_k)$, and $2\delta_1 = \min \mid \xi_i - \xi_j \mid$ for $i \neq j$. Since $w_1$ and $w_2$ are open, cubes with sides parallel to the coordinate hyperplanes ($x_j =$ constant) and edges of length $2\delta_2$ may be centered on the points $E_0'$ so that each cube is entirely in $w_1$ or entirely in $w_2$, by choosing $\delta_2$ sufficiently small. Choose $\delta$ so that $\delta > 0$, $\delta < \delta_1$, $\delta < \delta_2$. The set $\{E_0'\}$ is a subset of the set $\{E_0''\}$ of $k^k$ points whose coordinates are in the set $\xi_1, \cdots, \xi_k$ allowing repetitions. For each point $E_0'' = (\xi_{i_1}, \cdots, \xi_{i_k})$ in $\{E_0''\}$ construct a cube $C_{i_1, \cdots, i_k}$ as above with center at $E_0''$ and edge $2\delta$. These cubes are disjoint. Let $f_i(x)$ be a p.d.f. such that the corresponding c.d.f. is in $\Omega_4$ and $f_i(x) = 0$ for $|x - \xi_i| > \delta$ ($i = 1, \cdots, k$). Define the p.d.f.

$$f^{(s)}(x) = s^{-1} \sum_{i=1}^{s} f_i(x) \qquad (s = 1, \cdots, k).$$

Then the corresponding c.d.f. $F^{(s)}$ is in $\Omega_4$. We have

$$\alpha = P(w \mid F^{(s)}) = \int_w \prod_{j=1}^{k} f^{(s)}(x_j)\, dW$$

$$= s^{-k} \int_w \sum_{i_1, \cdots, i_k = 1}^{s} f_{i_1}(x_1) \cdots f_{i_k}(x_k)\, dW,$$

where $dW = dx_1 \cdots dx_k$. Bring the last summation sign outside the integral sign, and note that $f_{i_1}(x_1) \cdots f_{i_k}(x_k) = 0$ outside $C_{i_1, \cdots, i_k}$. Then

$$\sum_{i_1, \cdots, i_k = 1}^{s} I_{i_1, \cdots, i_k} = s^k \alpha, \tag{3}$$

where

$$I_{i_1, \cdots, i_k} = \int_{w \cap C_{i_1, \cdots, i_k}} f_{i_1}(x_1) \cdots f_{i_k}(x_k)\, dW. \tag{4}$$

Our argument depends on certain sums of $I_{i_1, \cdots, i_k}$ having the property that the sum is equal to $\alpha$ times the number of terms in the sum. In order to save space we shall say that if $\Sigma$ is such a sum, then $\Sigma \in R$, $R$ being the class of such sums. Clearly all sums (3) are in $R$. Let $\{S_{r\nu}\}$ be the subsets of $r$ ($r = 1, \cdots, k$) different integers in the set $1, 2, \cdots, k$ ($\nu = 1, \cdots, {}_kC_r$), and let $\Sigma_{r\nu}$ be the sum of all $I_{i_1, \cdots, i_k}$ for which the index $i_1, \cdots, i_k$ consists only of integers in $S_{r\nu}$ and such that *all* the integers of $S_{r\nu}$ appear in the index. We wish to prove that $\Sigma_{k1}$, the sum of $I$ for cubes centered on the points of $\{E_0'\}$, is in $R$. To accomplish this we make an induction on $r$: If we assume all $\Sigma_{r\nu} \in R$ for $r < s$, then we can show all $\Sigma_{s\mu} \in R$ ($s = 2, \cdots, k$). No generality is lost in taking $S_{s\mu}$ as the set of integers $1, 2, \cdots, s$. Now consider the left member of (3). Some thought will show[5] that it may be broken down into $\Sigma_{s\mu}$ plus a sum of $\Sigma_{r\nu}$ where $r < s$. But the left member of (3) is in $R$, and by hypothesis so are all $\Sigma_{r\nu}$ with $r < s$. It follows that $\Sigma_{s\mu}$ is also in $R$. To see that $\Sigma_{1\nu} \in R$ ($\nu = 1, \cdots, k$), let

[5] To illustrate the reasoning, suppose $s = 4$. If $S_{\sigma\tau}$ is the set of (different) integers $a, b, \cdots, h$, denote $\Sigma_{\sigma\tau}$ by <$a, b, \cdots, h$>, that is, <$a, b, \cdots, h$> is the sum of all $I$ whose indices contain $a, b, \cdots, h$ and no other integers. Then the right member of (3) contains terms from <1, 2, 3, 4>; <1, 2, 3>, <1, 2, 4>, <1, 3, 4>, <2, 3, 4>; <1, 2>, <1, 3>, <1, 4>, <2, 3>, <2, 4>, <3, 4>; <1>, <2>, <3>, <4>. Every term of the right member of (3) is in one of these sums < >. No term can appear in 2 sums < >. Every term of each sum < > appears in the right member of (3). Thus the right member is the sum of all sums < > listed above, and by hypothesis, all but the first sum < > are in $R$.

$S_{1\nu}$ be $\nu$ and note that $\Sigma_{1\nu}$ consists only of $I_{\nu,\nu,\ldots,\nu}$. Putting $s = 1$ in (3) we have $I_{1,1,\ldots,1} = \alpha$, and likewise $\Sigma_{1\nu} = I_{\nu,\nu,\ldots,\nu} = \alpha$. Thus $\Sigma_{1\nu} \epsilon R$.

We have at this stage that $\Sigma_{k1} = k!\alpha$. But as we already noted, of the cubes $C$ associated with the integrals $I$ in the sum $\Sigma_{k1}$, $M_0$ are entirely inside $w_1$ and $k! - M_0$ entirely outside $w_1$. For the set of $M_0$ terms in $\Sigma_{k1}$ corresponding to the cubes $C$ in $w_1$ the region of integration $w \cap C$ in (4) is actually $C$, and for the remaining set of terms in $\Sigma_{k1}$ the region of integration is the empty set. Furthermore if $w \cap C = C$ in (4), the corresponding $I$ is unity. Hence $\Sigma_{k1} = M_0 = k!\alpha$, $\alpha = M_0/k!$. If we now repeated the process with any other point $E_1 \epsilon w_1$ instead of $E_0$, and let $M_1$ be the number of points of $\{E_1'\}$ in $w_1$, we would get $\alpha = M_1/k!$. Therefore $M_1 = M_0$. From $0 < \alpha < 1$, we conclude $0 < M_0 < k!$. Thus $w_1$ has the structure $S$.

The exceptional null set allowed for in the statement of Theorem 3 entered the proof when we removed $w \cap H'$ from $w$. Had we assumed that the boundary $B \epsilon N_2$, then the exceptional set would be in $N_2$. As a corollary to the reasoning used in the proof we thus get

COROLLARY 3: *If the boundary of $w$ is in $N_2$, a necessary condition that $w$ have the property $\pi_4$ is that $w$ have the structure $S$ except on a subset in $N_2$.*

Finally, because of (2), any sufficient (necessary) condition for $w$ to have the property $\pi_i$ is sufficient (necessary) for $w$ to have the property $\pi_j$ if $j > i$ $(j < i)$. Hence we may replace $\pi_2$ in Theorem 2 and Corollary 1 by $\pi_3$ or $\pi_4$, $\pi_3$ in Corollary 2 by $\pi_4$, $\pi_4$ in Theorem 3 and Corollary 3 by $\pi_3$ or $\pi_2$. This yields

COROLLARY 4: *If the boundary of $w$ is a null set, a necessary and sufficient condition that $w$ have the property $\pi_3$ (or $\pi_4$) is that it have the structure $S$ except on a null set.*

COROLLARY 5: *If the boundary of $w$ is a region in $N_2$, a necessary and sufficient condition that $w$ have the property $\pi_2$ (or $\pi_3$ or $\pi_4$) is that it have the structure $S$ except on a subset in $N_2$.*

**3. Remarks.** Wald and Wolfowitz [6, 8] in their work on the problem of two samples for the case $F \epsilon \Omega_2$ have imposed the following restriction on any statistic used to test the null hypothesis: The statistic must be a function of $V$ only, where the sequence $V$ of $k$ elements is formed as follows: Rank the $X_j$ of the sample in ascending order of magnitude (ignoring cases where two $X_j$ are equal), and if the $i$-th element in this rank order is a $Y$ put the $i$-th element of $V$ equal to zero, else unity. This means that the resulting critical region always consists of the union of $s$ of the regions $u_i$ defined in section 2, where $s$ is a multiple of $m!n!$. The results of our section 2 show that this restriction is not necessary if all we require is that $Pr\{E \epsilon w\}$, where $w$ is the critical region and $E$ the sample point, be the same constant $\alpha$ whenever the null hypothesis is true. In fact a valid (but probably not very efficient) solution of the problem of two samples has been proposed by Pitman [3] in which the statistic is not a function of $V$ only.

Putting further requirements on the critical region will lead to a more restricted class than the class of regions having essentially the structure $S$. For instance, from section 2 it follows that the significance level $\alpha$ can be any of the values $i/k!$ $(i = 1, \cdots, k! - 1)$. But if we lay down a symmetry condition to the effect that if $(y_1, \cdots, y_m, z_1, \cdots, z_n)$ is in $w$, all points obtainable by permuting the $y$'s among themselves and the $z$'s among themselves be in $w$, then $\alpha$ must be a multiple of $m!n!/k!$. Again, if we impose the condition that any statistic $T(X_1, \cdots, X_k)$ used to test the null hypothesis remain invariant when all the $X_j$ are subjected to the same topological transformation of the real line onto itself, then Wald and Wolfowitz [6] have shown that $T$ must be a function of $V$ only, so that $w$ has the special structure described above. It would seem desirable when the subject of statistical inference in the non-parametric case may be entering a stage of rapid development, to be clear about the assumptions necessary to restrict the critical region to a particular class.

In concluding these remarks, we quote with the kind permission of Dr. Wolfowitz, from some correspondence with the writer. Important work has been done on non-parametric tests under the restriction that the statistic used be invariant under topological transformation. The following statement as to *why* this restriction might be imposed will therefore interest the reader: " $\cdots$ there are arguments pro and con $\cdots$ *Pro:* If the statistic be not invariant, this could happen: Two scientists working on the same problem and having the *same* observations to interpret might come to opposite conclusions if one used one scale of measurement and the other used a monotone function of that scale. *Con:* The criterion of topologic invariance of the statistic is a restriction on our freedom. Furthermore it cannot be imposed except in the univariate case ([8], p. 270)."

## REFERENCES

[1] W. J. DIXON, *Annals of Math. Stat.*, Vol. 11 (1940), pp. 199–204.
[2] A. KOLMOGOROFF, *Annals of Math. Stat.*, Vol. 12 (1941), pp. 461–463.
[3] E. J. G. PITMAN, *J. Roy. Stat. Soc. Suppl.*, Vol. 4 (1937), pp. 117–130.
[4] S. SAKS, *Theory of the Integral*, Warsaw, 1937.
[5] A. WALD and J. WOLFOWITZ, *Annals of Math. Stat.*, Vol. 10 (1939), pp. 105–118.
[6] A. WALD and J. WOLFOWITZ, *Annals of Math. Stat.*, Vol. 11 (1940), pp. 147–162.
[7] S. S. WILKS, *Annals of Math. Stat.*, Vol. 12 (1941), pp. 91–96.
[8] J. WOLFOWITZ, *Annals of Math. Stat.*, Vol. 13 (1942), pp. 247–279.
[9] W. FELLER, *Stat. Res. Mem.*, Vol. 2 (1938), pp. 107–125.
[10] R. A. FISHER, *Statistical Methods for Research Workers*, section 24, example 19; *The Design of Experiments*, section 21; *J. Roy. Anthrop. Soc.*, Vol. 66 (1936), pp. 57–63.
[11] A. KOLMOGOROFF, *Gior. Ist. Ital. Attuari*, Vol. 4 (1933), pp. 83–91.
[12] J. NEYMAN, *J. Roy. Stat. Soc.*, Vol. 105 (1942), pp. 311–312.