# 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.