SIAM J. APPL. MATH.
Vol. 38, No. 3, June 1980

0036-1399/80/3803-0009 $01.00/0

DETECTION OF ABNORMAL BEHAVIOR VIA NONPARAMETRIC ESTIMATION OF THE SUPPORT*

LUC DEVROYE† AND GARY L. WISE‡

Abstract. In this paper two problems are considered, both involving the nonparametric estimation of the support of a random vector from a sequence of independent identically distributed observations. In the first problem, after observing n independent random vectors with a common unknown distribution μ, we are given one new measurement and we wish to know whether or not it belongs to the support of μ. In the second problem, after observing the n independent random vectors with a common unknown distribution μ, we then observe n additional independent random vectors with a common unknown distribution ν. In this case we wish to know whether or not the support of ν is completely contained within the support of μ. Decision schemes are presented and then convergence properties are established.

1. Introduction. A problem of increasing significance to engineers concerns the detection of abnormal or faulty behavior of a system, plant, or machine. We assume that we have observed the system in normal operation and that we have taken measurements of the normal behavior. A measurement is assumed to be an $\mathbb{R}^d$-valued random vector. All random vectors in this paper are defined on a probability space $(\Omega, \mathcal{F}, P)$. The randomness may be due to measurement noise, parasitic effects, or random inputs. Thus the measurements are given by $X_1, X_2, \cdots, X_n$, a sequence of $\mathbb{R}^d$-valued random vectors which we assume are independent with a common unknown probability measure μ defined on the Borel sets of $\mathbb{R}^d$.

Classically, the assumption is made that one has access at the present time to m independent observations $X'_1, X'_2, \cdots, X'_m$ with common probability measure ν, and the system is said to behave differently, or abnormally, if $\nu \neq \mu$. To detect such a change in distribution, several tests have been proposed (Kolmogorov [5], Smirnov [8], Cramér [1], von Mises [9], Lehmann [6], Renyi [7], and Wald and Wolfowitz [10]). For a survey of tests of this type, we refer to the book of Hajek and Sidak [3, pp. 90–94] and the early survey article of Darling [2].

The first problem we treat in this paper is concerned with taking one new observation. For economic reasons, lack of time, or practical limitations, only one new observation X can be made and there is no hope to recover or approximate ν as with the large sample $X'_1, X'_2, \cdots X'_m$. Regardless of ν, we say that the system behaves abnormally if $X \notin S$, the support of μ. In several practical applications, the complement S^c of S can be considered as a danger area because under normal behavior (with probability measure μ) the probability that some of the X_i take values in S^c is zero. We thus have reduced the problem to one of estimating the support S from $X_1, X_2, \cdots, X_n$. This problem is considered in the next section. Then in § 3 we introduce a counterexample to illustrate the necessity of certain conditions imposed in § 2.

The second problem that we consider is concerned with taking n new measurements which are independent with common unknown probability measure ν. In this case we assume that the system might have changed, but we are concerned with whether or not the system might exhibit abnormal behavior. We assume that the system still functions normally if the support of ν is contained within S. This problem is treated in § 4.

* Received by the editors February 27, 1978, and in revised form June 14, 1979. This research was supported in part by the Air Force Office of Scientific Research, under Grant AFOSR-76-3062.

† School of Computer Science, McGill University, Montreal, Canada H3A 2K6.

‡ Department of Electrical Engineering, University of Texas, Austin, Texas 78712.

2. Detection based on a single new observation. We assume that we have one new observation X, and we want to know whether or not X belongs to S, the support of μ. Therefore, we will estimate S from $X_1, X_2, \cdots, X_n$, and base our decision upon this estimate. The obvious estimate is

$$S_n = \bigcup_{i=1}^{n} A(X_i, \rho_n), \tag{1}$$

where ρ_n is a number depending only upon n and $A(x, a)$ is the closed sphere centered at x with radius $a \geqq 0$ given by

$$A(x, a) = \{y : y \in \mathbb{R}^d \text{ and } \|x - y\| \leqq a\},$$

where $\|\cdot\|$ denotes the $L_p (p \geqq 1)$ norm on $\mathbb{R}^d$. The decision rule is simple:

$$\text{Decide } X \notin S \text{ if } X \notin S_n.$$

$$\text{Decide } X \in S \text{ if } X \in S_n.$$

The probability of making an error, given $X_1, X_2, \cdots, X_n$, is

$$L_n = P\{X \in S \Delta S_n | X_1, X_2, \cdots, X_n\}$$
$$= \nu(S \Delta S_n).$$

The detection procedure is said to be consistent if $L_n \xrightarrow{n} 0$ in probability. It is called strongly consistent if $L_n \xrightarrow{n} 0$ with probability one (wp1). In this section we give two theorems; the first establishes consistency, and the second establishes strong consistency.

THEOREM 1. *Let μ be any probability measure on the Borel sets of $\mathbb{R}^d$, let ν be* any *probability measure on the Borel sets of $\mathbb{R}^d$ whose restriction to S is absolutely continuous with respect to μ, and let the positive numbers ρ_n satisfy*

$$\rho_n \xrightarrow{n} 0 \tag{2}$$

and

$$n\rho_n^d \xrightarrow{n} \infty. \tag{3}$$

Then $L_n \xrightarrow{n} 0$ in probability.

Proof. Let M and α be positive numbers and let

$$B_n = [-M, M]^d \cap \{x : \mu(A(x, \rho_n)) \geqq \alpha \rho_n^d\}.$$

Let ε and δ be arbitrary positive numbers. Noting that

$$L_n = \int_{S_n \cap S^c} \nu(dx) + \int_{S_n^c \cap S} \nu(dx),$$

we have that

$$P\{L_n \geqq 3\varepsilon\} \leqq \frac{1}{\varepsilon} \nu(S \cap B_n^c) + P\left\{\int_{S_n \cap S^c} \nu(dx) \geqq \varepsilon\right\} + P\left\{\int_{B_n \cap S_n^c \cap S} \nu(dx) \geqq \varepsilon\right\}. \tag{4}$$

Now note that

$$\frac{1}{\varepsilon} \nu(S \cap B_n^c) \leqq \frac{1}{\varepsilon} \nu\{([-M, M]^d)^c\} + \frac{1}{\varepsilon} \int_{S \cap C_n} \nu(dx),$$

where

$$C_n = \{x : \mu(A(x, \rho_n)) < \alpha \rho_n^d\}.$$

If M is sufficiently large, we have that

$$\frac{1}{\varepsilon}\nu(S \cap B_n^c) \leq \frac{\delta}{6} + \frac{1}{\varepsilon}\int_{S \cap C_n} \nu(dx).$$

Now we show that as $n \to \infty$, the set C_n has μ-measure smaller than an arbitrary positive number γ if α is sufficiently small. Notice that, for K sufficiently large, we have that

$$\mu(C_n) \leq \mu(C_n \cap [-K, K]^d) + \frac{\gamma}{2}.$$

Let n' be such that for all $n > n'$, $\rho_n \leq 1$; and assume that $n > n'$. Now find N points $x_1, x_2, \cdots, x_N$ in $[-K, K]^d$ such that for every x in $[-K, K]^d$, $\|x - x_i\| \leq 1$ for some i. Then we have that

$$\begin{aligned}\int_{C_n} \mu(dx) &\leq \sum_{i=1}^{N} \int_{C_n \cap A(x_i, 1)} \mu(dx) + \frac{\gamma}{2} \\ &\leq N\alpha + \frac{\gamma}{2} \\ &< \gamma\end{aligned}$$

for α sufficiently small. Since ν restricticted to S is absolutely continuous with respect to μ, we know that we can pick γ so small that $\mu(C_n) < \gamma$ implies that $(1/\varepsilon)\nu(SC_n) < \delta/6$; and thus the first term in (4) can be made arbitrarily small if $n > n'$.

By the Lebesgue dominated convergence theorem, we have that

$$\nu(T_n \cap S^c) \to 0 \quad \text{as } \rho_n \to 0,$$

where

$$T_n = \bigcup_{x \in S} A(x, \rho_n),$$

because S, the support of μ, is a closed set. (Recall that the support of μ is the set of all x such that $\mu(A(x, \varepsilon)) > 0$ for all $\varepsilon > 0$; equivalently, it is the smallest closed set with μ-measure one.) Since $X_i \in S$ wp 1, we have that

$$\int_{S_n \cap S^c} \nu(dx) \leq \int_{T_n \cap S^c} \nu(dx) = \nu(T_n \cap S^c) \xrightarrow{n} 0,$$

and thus the second term in (4) goes to zero as $n \to \infty$.

Next, we need only show that

$$P\{\nu(B_n S_n^c S) \geq \varepsilon\} < \frac{\delta}{3}$$

for n sufficiently large. By Markov's inequality, and writing I to denote the indicator function, we have

$$
\begin{aligned}
P\{\nu(B_nS_n^cS) \geqq \varepsilon\} &\leqq \frac{1}{\varepsilon} E\{\nu(B_nS_n^cS)\} \\
&= \frac{1}{\varepsilon} E\left[I_{\{B_n \cap S\}} P\left\{ \bigcap_{i=1}^{n} (\|X_i - X\| > \rho_n) | X \right\} \right] \\
&= \frac{1}{\varepsilon} E\{I_{\{B_n \cap S\}} [1 - \mu(A(X, \rho_n))]^n\} \\
&\leqq \frac{1}{\varepsilon} E\{I_{\{B_n \cap S\}} \exp[-n\mu(A(X, \rho_n))]\} \\
&\leqq \frac{1}{\varepsilon} E[I_{\{B_n \cap S\}} \exp(-n\alpha\rho_n^d)] \\
&\leqq \frac{1}{\varepsilon} \exp(-\alpha n \rho_n^d) \\
&\xrightarrow{n} 0. \quad \text{QED.}
\end{aligned}
$$

The strength of Theorem 1 is that it holds for *all* probability measures μ and a large class of probability measures ν. For instance, the condition of the theorem holds if both μ and ν are absolutely continuous with respect to Lebesgue measure.

Now we present a theorem which establishes the strong consistency of the detection procedure.

THEOREM 2. *Let μ be any probability measure on the Borel sets of $\mathbb{R}^d$, let ν be any probability measure on the Borel sets of $\mathbb{R}^d$ whose restriction to S is absolutely continuous with respect to μ, and let the positive number ρ_n satisfy*

$$\rho_n \xrightarrow{n} 0$$

and

$$\sum_{n=1}^{\infty} \exp(-\alpha n \rho_n^d) < \infty \quad \textit{for all } \alpha > 0. \tag{5}$$

Then $L_n \xrightarrow{n} 0$ wp1.

Proof. Using the notation in the proof of Theorem 1, notice that

$$
\begin{aligned}
P\left\{ \bigcup_{n \geqq k} (L_n \geqq 3\varepsilon) \right\} \leqq \frac{1}{\varepsilon} \nu(S \cap B_n^c) + P\left\{ \bigcup_{n \geqq k} [\nu(S_n \cap S^c) \geqq \varepsilon] \right\} \\
+ P\left\{ \bigcup_{n \geqq k} [\nu(B_n \cap S \cap S_n^c) \geqq \varepsilon] \right\}.
\end{aligned}
$$

The first term in the above upper bound is smaller than $\delta/3$ as shown before. The second term is upper bounded by

$$P\left\{ \bigcup_{n \geqq k} [\nu(T_n \cap S^c) \geqq \varepsilon] \right\}.$$

Let

$$n_k = \min\{r : r \geqq k \text{ and } \rho_r \geqq \rho_n \text{ for all } n \geqq k\}.$$

Then we have that

$$P\left\{\bigcup_{n\geqq k}[\nu(T_n\cap S^c)\geqq\varepsilon]\right\}\leqq P\{\nu(T_{n_k}\cap S^c)\geqq\varepsilon\}\xrightarrow{k}0,$$

by an argument as in the proof to Theorem 1. Finally, by (5) and an inequality developed in the proof to Theorem 1, we have that

$$P\left\{\bigcup_{n\geqq k}[\nu(B_n\cap S\cap S_n^c)\geqq\varepsilon]\right\}\leqq\sum_{n\geqq k}P\{\nu(B_n\cap S\cap S_n^c)\geqq\varepsilon\}$$

$$\leqq\sum_{n\geqq k}\frac{1}{\varepsilon}\exp(-\alpha n\rho_n^d)\xrightarrow{k}0.$$

Theorem 2 follows by the arbitrariness of ε. Q.E.D.

Notice that condition (5) is satisfied if

$$\frac{n\rho_n^d}{\log n}\xrightarrow{n}\infty.$$

The converse of this statement is not always true; however, the monotonicity of ρ_n is sufficient for the converse.

3. A counterexample. In the preceding section, the area controlled by the spheres $A(X_i,\rho_n)$ is, roughly speaking, proportional to $n\rho_n^d$. Condition (5) is natural because it allows this area to grow in case S has a very large (or infinite) Lebesgue measure. What puts a restriction on the detector is the condition that the restriction of ν to S is absolutely continuous with respect to μ. We now show that given any sequence $\{\rho_n\}$ satisfying (2) there exist μ and ν such that the procedure is not consistent. Thus adding any condition on $\{\rho_n\}$ besides (2) will not enable us to remove the condition on μ and ν.

Let $\rho_n\xrightarrow{n}0$ and let f be a density function on $\mathbb{R}^d$ such that

$$n\int_{A(\theta,\rho_n)}f(x)\,dx\xrightarrow{n}0,$$

where θ denotes the origin. Let μ have density f and let ν be atomic at the origin. Clearly,

$$P\{L_n=1\}\geqq\left[\int_{(A(\theta,\rho_n))^c}f(x)\,dx\right]^n$$

$$=\left[1-\int_{A(\theta,\rho_n)}f(x)\,dx\right]^n$$

$$\geqq\exp\left[\frac{-n\int_{A(\theta,\rho_n)}f(x)\,dx}{1-\int_{A(\theta,\rho_n)}f(x)\,dx}\right]\xrightarrow{n}1.$$

To construct such a function f, let $\bar\rho_n=1/n+\sup_{k\geqq n}\rho_k$, and let

$$f(x)=\sum_{i=1}^{\infty}\alpha_i I_{\{D_i\}}(x),$$

where

$$D_i=A\left(y_i,\frac{\bar\rho_i-\bar\rho_{i+1}}{2}\right),$$

y_i is a point in $\mathbb{R}^d$ whose first coordinate is $\rho_i + (\bar{\rho}_i - \bar{\rho}_{i+1})/2$ and all other coordinates are zero and

$$\alpha_i = \left[\frac{1}{i^2} - \frac{1}{(i+1)^2}\right]\left[\int_{D_i} dx\right]^{-1}.$$

Clearly, f is a density and

$$n \int_{A(\theta,\rho_n)} f(x)\, dx \leqq n \sum_{i=n}^{\infty} \frac{1}{i^2} - \frac{1}{(i+1)^2} = \frac{1}{n} \xrightarrow{n} 0.$$

4. A detection problem. Assume that we now have two sequences, $X_1, X_2, \cdots, X_n$ and $Y_1, Y_2, \cdots, Y_n$, of independent $\mathbb{R}^d$-valued random vectors. The X_i have probability measure μ and the Y_i have probability measure ν, both unknown. For the sake of simplicity, we have assumed that the sequences are of equal length. The X_i again correspond to normal behavior, and we say that the Y_i are abnormally distributed if with positive probability Y_1 takes values outside S, the support of μ; that is, if $\nu(S^c) > 0$. Notice that this problem is quite different from the one of the detection of a change in distribution function (the problem of detecting whether $\nu = \mu$ or not). We propose the following detection procedure. Let us construct S_n as in (1) from the X_i. Define

$$Z_i = I_{\{Y_i \notin S_n\}}, \qquad 1 \leqq i \leqq n,$$

and let

$$N_n = \sum_{i=1}^{n} Z_i$$

be the number of Y_i falling outside S_n. We decide that $\nu(S^c) = 0$ if $N_n < k_n$ and that $\nu(S^c) > 0$ is $N_n \geqq k_n$, where $\{k_n\}$ is a sequence of integers satisfying

$$k_n \xrightarrow{n} \infty$$

and

$$\frac{k_n}{n} \xrightarrow{n} 0.$$

Formally, we put

$$D_n = 1 \quad \text{if } N_n \geqq k_n,$$
$$D_n = 0 \quad \text{if } N_n < k_n.$$

The criteria of the goodness of the detection procedure now are

$$\tilde{L}_n = P\{D_n = 1\} \text{ under the hypothesis that } \nu(S^c) = 0$$

and

$$\hat{L}_{n,\delta} = P\{D_n = 0\} \text{ under the hypothesis that } \nu(S^c) = \delta > 0.$$

In the remainder of this section we introduce conditions insuring that $\tilde{L}_n$ and $\hat{L}_{n,\delta}$ tend to zero as n grows large.

THEOREM 3. *If* $L_n \xrightarrow{n} 0$ *in probability, then* $\hat{L}_{n,\delta} \to 0$.

Proof. Notice that

$$\frac{E\{N_n|X_1, X_2, \cdots, X_n\}}{n} = \int_{S_n^c} \nu(dx) = \nu(S^c) + \nu(S_n^c \cap S) - \nu(S_n \cap S^c)$$
$$\geqq \delta - \nu(S_n \cap S^c)$$
$$\geqq \delta - L_n.$$

Let n' be such that for all $n > n'$, $k_n/n < \delta/3$, and consider $n > n'$. Then we have that

$$P\{D_n = 0\} = P\{N_n < k_n\}$$
$$= E\left\{P\left\{\frac{N_n - E\{N_n|X_1, X_2, \cdots, X_n\}}{n} < \frac{k_n - E\{N_n|X_1, X_2, \cdots, X_n\}}{n} \Big| X_1, X_2, \cdots, X_n\right\}\right\}$$
$$\leqq P\{L_n \geqq \delta/3\} + E\left\{P\left\{\frac{N_n - E\{N_n|X_1, X_2, \cdots, X_n\}}{n} < -\frac{\delta}{3} \Big| X_1, X_2, \cdots, X_n\right\}\right\}.$$

Notice that, conditioned on $X_1, X_2, \cdots, X_n$, the Z_i are independent identically distributed $\{0, 1\}$-valued random variables. Using the inequality of Hoeffding [4], we get that

$$P\{D_n = 0\} \leqq P\{L_n \geqq \delta/3\} + \exp\left[-2n\left(\frac{\delta}{3}\right)^2\right] \xrightarrow{n} 0. \quad \text{Q.E.D.}$$

COROLLARY. *Assume that* (2) *and* (3) *hold and that* μ *and* ν *are probability measures on the Borel sets of* $\mathbb{R}^d$ *such that the restriction of* ν *to* S *is absolutely continuous with respect to* μ. *If* $\nu(S^c) = \delta > 0$, *then* $\hat{L}_{n,\delta} \xrightarrow{n} 0$.

THEOREM 4. *If* $k_n \xrightarrow{n} \infty$, $\sup_n \rho_n < \infty$,

$$\frac{n\rho_n^{2d}}{\log n} \xrightarrow{n} \infty \quad \text{and} \quad \frac{k_n}{n\rho_n^d} \geqq b > 0,$$

if ν *has compact support completely contained in* S, *the support of* μ, *and if there exists a finite constant* K *with*

$$(6) \qquad \nu(A(x, \alpha)) \leqq K\mu(A(x, \alpha)) \quad \text{for all } x \in S \quad \text{and} \quad \alpha > 0,$$

then $\tilde{L}_n \xrightarrow{n} 0$.

Proof. Pick M so that $\nu([-M, M]^d) = 1$ and define the sets B_n and C_n by

$$B_n = [-M, M]^d \cap \{x : \mu(A(x, \rho_n)) \geqq \beta\rho_n^{2d}\},$$
$$C_n = [-M, M]^d \cap \{x : \mu(A(x, \rho_n)) < \beta\rho_n^{2d}\},$$

where β is a positive constant.

Let $x_1, \cdots, x_{m_n}$ be picked such that for all x in the support of ν, $\|x - x_i\| \leqq \rho_n/2$ for some i. Let γ be a constant depending upon p, M, and d, such that $m_n = [1 + (\gamma/\rho_n)]^d$.

Then we have that

$$\begin{aligned} \nu(SB_n^c) &\leqq \nu(C_n) \\ &\leqq \sum_{i=1}^{m_n} \nu\Big(C_n A\Big(x_i, \frac{\rho_n}{2}\Big)\Big) \\ &\leqq K \sum_{i=1}^{m_n} \mu\Big(C_n A\Big(x_i, \frac{\rho_n}{2}\Big)\Big). \end{aligned} \tag{7}$$

Since $x \in C_n A(x_i, \rho_n/2)$ implies that $C_n A(x_i, \rho_n/2) \subset A(x, \rho_n)$, it follows by the definition of C_n that (7) is further upper bounded by

$$\Big(1+\frac{\gamma}{\rho_n}\Big)^d K\beta\rho_n^{2d}$$

which by choice of

$$\beta < \frac{b}{3K2^{d-1}(\gamma^d + \sup_n \rho_n^d)}$$

is smaller than $b\rho_n^d/3$. Now,

$$\begin{aligned} \frac{E\{N_n|X_1, \cdots, X_n\}}{n} &= \nu(S_n^c S) \\ &\leqq \nu(SB_n^c) + \nu(S_n^c SB_n), \end{aligned}$$

and

$$\begin{aligned} P\{D_n = 1\} &= P\{N_n \geqq k_n\} \\ &\leqq E\Big\{P\Big\{\frac{N_n - E\{N_n|X_1, \cdots, X_n\}}{n} \geqq \frac{k_n}{n} - \nu(SB_n^c) - \nu(S_n^c SB_n)|X_1, \cdots, X_n\Big\}\Big\} \\ &\leqq E\Big\{P\Big\{\frac{N_n - E\{N_n|X_1, \cdots, X_n\}}{n} \geqq \frac{k_n}{n} - \frac{2b}{3}\rho_n^d|X_1, \ldots, X_n\Big\}\Big\} \\ &\quad + P\Big\{\nu(S_n^c SB_n) \geqq \frac{b\rho_n^d}{3}\Big\} \\ &\leqq E\Big\{P\Big\{\frac{N_n - E\{N_n|X_1, \cdots, X_n\}}{n} \geqq \frac{b\rho_n^d}{3}\Big|X_1, \cdots, X_n\Big\}\Big\} \\ &\quad + \frac{3}{b\rho_n^d} \exp(-\beta n \rho_n^{2d}) \\ &\leqq \exp\Big[-2n\Big(\frac{b\rho_n^d}{3}\Big)^2\Big] + \frac{3}{b\rho_n^d}\exp(-\beta n\rho_n^{2d}). \end{aligned}$$

Here we used Hoeffding's inequality [4] and an inequality from the proof of Theorem 1. The first term on the right hand side of the above bound obviously goes to zero as $n \to \infty$. To see that the second term does also, let

$$\frac{n\rho_n^{2d}}{\log n} = g(n),$$

and notice that

$$\frac{\exp(-\beta n\rho_n^{2d})}{\rho_n^d} = \frac{\exp\left(\log n\left[\frac{1}{2\log e} - \beta g(n)\right]\right)}{(g(n)\log n)^{1/2}} \xrightarrow{n} 0 \quad \text{Q.E.D.}$$

COROLLARY. *Let $k_n = bn\rho_n^d$ for some $b > 0$, let $\rho_n \xrightarrow{n} 0$, and let $n\rho_n^{2d}/\log n \xrightarrow{n} \infty$. If condition* (6) *holds and if ν has compact support completely contained in S, then $\tilde{L}_n \xrightarrow{n} 0$.*

REFERENCES

[1] H. CRAMÉR, *On the composition of elementary errors*, Skand. Aktuarietidskr., 11 (1928), pp. 13–74, pp. 141–180.

[2] D. A. DARLING, *The Kolmogorov–Smirnov, Cramér–von Mises tests*. Ann. Math. Statist., 28 (1957), pp. 823–838.

[3] J. HAJEK AND Z. SIDAK, *Theory of Rank Tests*, Academic Press, New York, 1967.

[4] W. HOEFFDING, *Probability inequalities for sums of bounded random variables*, J. Amer. Statist. Assoc., 58 (1963), pp. 13–30.

[5] A. N. KOLMOGOROV, *Sulla determinizione empirica di una legge di distribuzione*, Giorn. Ist. Ital. Attuari, 4 (1933), pp. 83–91.

[6] E. H. LEHMANN, *Consistency and unbiasedness of certain nonparametric tests*, Ann. Math. Statist., 22 (1951), pp. 165–179.

[7] A. RENYI, *Neue Kriterien zum Vergleich zweier Stichproben*, Magyar Tud. Akad. Alkalm. Mat. Int. Közl., 2 (1953), pp. 243–265.

[8] N. V. SMIRNOV, *Estimate of deviation between empirical distribution functions in two independent samples*, Bull. Moscow Univ., 2 (1939), pp. 3–16.

[9] R. VON MISES, *Wahrscheinlichkeitsrechnung*, Leipzig, Wien, 1931.

[10] A. WALD AND J. WOLFOWITZ, *On a test whether two samples are from the same population*, Ann. Math. Statist., 11 (1940), pp. 147–162.