ISSN 8756-6990, Optoelectronics, Instrumentation and Data Processing, 2015, Vol. 51, No. 2, pp. 110–119. © Allerton Press, Inc., 2015.
Original Russian Text © G.I. Salov, 2015, published in Avtometriya, 2015, Vol. 51, No. 2, pp. 11–22.

New Nonparametric Statistical Test for Problems with Three Samples, which is More Effective than the Whitney Test

G. I. Salov

*Institute of Computational Mathematics and Mathematical Geophysics,
Siberian Branch, Russian Academy of Sciences,
pr. Akademika Lavrent'eva 6, Novosibirsk, 630090 Russia
E-mail: sgi@ooi.sscc.ru*

Received April 17, 2014

Abstract—A new nonparametric statistical test for testing the hypothesis of homogeneity of three samples against an alternative hypothesis, which implies that random variables of one of these samples tend to be stochastically greater than random variables of each of the other two samples separately, is proposed. The known Whitney test is equivalent to a particular case of the new test. Application of these tests in a real difficult problem of detection of a specified extended object in a noisy image in a situation with possible appearance of another "interfering" object of a greater size is considered.

Keywords: three samples, homogeneity test, nonparametric test, detection of objects, noisy image.

DOI: 10.3103/S8756699015020028

INTRODUCTION

The proposed study continues those described in [1–3], though this paper can be also read independently. In this paper, we consider a more complicated (generalized) example of object detection discussed in [1].

EXAMPLE OF A REAL DIFFICULT PROBLEM

As previously, let us consider a noisy image. Let us now assume that, in contrast to the example in [1], the observer's field of vision may include 1) the entire specified important extended object, 2) some part of the neighborhood of the contour of some object that has a greater size, but is not of interest for the observer ("interfering" object), or 3) neither of the objects. An important object for the observer is the specified object. The appearance of the interfering object can lead to the most undesirable false (MUF) "alarm." This fact should be taken into account.

Let $X_1, \ldots, X_m$ be a set of independent results of observations (measurements) performed at m points of the domain of the possible location of the important object, and let two more sets of independent observations $Y_1, \ldots, Y_n$ and $Z_1, \ldots, Z_n$ be taken for detection of this object in the case of its presence on both sides of this domain (symmetric with respect to the greatest midline of the domain). Moreover, let the observer know from the "physical" meaning and/or past experience the following: if neither the important nor interfering object appears in the field of vision during the observation period (hypothesis H_0), then $X_1, \ldots, X_m, Y_1, \ldots, Y_n$, and $Z_1, \ldots, Z_n$ can be considered as stochastically independent random variables with an identical continuous probability distribution function, which can be assumed to be $F(x)$ ("usual" homogeneity). If only the important object appears in the field of vision during the observation period (hypothesis H_1), then each of the variables X_i has another continuous probability distribution function, e.g., $G \leq F$, $G \not\equiv F$, i.e., the variables $X_1, \ldots, X_m$ are stochastically greater than $Y_1, \ldots, Y_n$ and $Z_1, \ldots, Z_n$ separately. Finally, if some part of the contour of the interfering object of a greater size is located in the field of vision during the observation period, then, in addition to X_i, each of the variables Y_j (or Z_j) has a different distribution function; let the latter coincide with the function $G(x)$ for simplicity. As in [1], the only information available for the observer about the functions $F(x)$ and $G(x)$ is the fact that they are continuous (this is a typical situation in practice).

The problem is to detect the presence of the important object on the basis of three independent sets (samples) $X_1, \ldots, X_m$, $Y_1, \ldots, Y_n$, and $Z_1, \ldots, Z_n$. It is necessary to indicate a test that would yield a desired result with the maximum (or close to maximum) probability and would be simultaneously little sensitive to the presence of the interfering object in the field of vision. The latter requirement, which has a contradictory character, makes this problem rather difficult. Among many known tests, the most suitable test for the problem considered here is the widely used nonparametric statistical test proposed by Whitney [4].

The goal of this paper is to propose a test that ensures a greater probability of object detection with a smaller (or equal) probability of the MUF alarm, i.e., a test that is more efficient than the Whitney test.

NEW TEST

The Whitney test (for brevity denoted by Wh) is based on the statistics U_1 and U_2 of two Mann–Whitney (MW) tests [5] and reads that the hypothesis H_0 about the homogeneity should be rejected in favor of H_1 if the following conditions are simultaneously satisfied:

$$U_1 = \sum_{i=1}^{m} \sum_{j=1}^{n} \mathbf{I}\{X_i > Y_j\} > C, \qquad U_2 = \sum_{i=1}^{m} \sum_{j=1}^{n} \mathbf{I}\{X_i > Z_j\} > C. \tag{1}$$

Hereinafter, $\mathbf{I}\{A\}$ is the function-indicator of an event A, which is equal to unity if the event A occurred and to zero otherwise, and the number C is chosen in accordance with the level of significance of the test specified in advance.

In [1], new statistics and a new nonparametric test based on these statistics and denoted by S_E were introduced. The results of the investigations in [1, 2] give grounds to assume that the S_E test can be more powerful than the MW test in many cases. Therefore, it is of interest to see what happens if two MW tests in the Whitney test are replaced by two S_E tests. It is desirable to avoid enhancement of the MUF alarm probability with increasing test power (detection probability).

For this purpose, we assume that $n = 2\nu$ is even and introduce the events ($i = 1, \ldots, m$, $j = 1, \ldots, \nu$)

$$E_{1ij}^{+} = \{X_i > \max(Y_j, Y_{\nu+j})\}, \quad E_{1ij}^{-} = \{X_i < \min(Y_j, Y_{\nu+j})\}, \quad E_{1ij}^{0} = \bar{E}_{1ij}^{+} \cap \bar{E}_{1ij}^{-},$$

$$E_{2ij}^{+} = \{X_i > \max(Z_j, Z_{\nu+j})\}, \quad E_{2ij}^{-} = \{X_i < \min(Z_j, Z_{\nu+j})\}, \quad E_{2ij}^{0} = \bar{E}_{2ij}^{+} \cap \bar{E}_{2ij}^{-},$$

and the statistics calculating the number of these events ($q = 1, 2$)

$$S_{Eq}^{+} = \sum_{i=1}^{m} \sum_{j=1}^{\nu} \mathbf{I}\{E_{qij}^{+}\}; \qquad S_{Eq}^{-} = \sum_{i=1}^{m} \sum_{j=1}^{\nu} \mathbf{I}\{E_{qij}^{-}\}; \qquad S_{Eq}^{0} = \sum_{i=1}^{m} \sum_{j=1}^{\nu} \mathbf{I}\{E_{qij}^{0}\}, \tag{2}$$

which take the values from 0 to $m\nu$ with the sum $S_{Eq}^{+} + S_{Eq}^{-} + S_{Eq}^{0} = m\nu$.

Using statistics (2), we obtain a new test

$$S_{Eq}^{+} > h(S_{Eq}^{0}), \quad q = 1, 2. \tag{3}$$

For brevity, it is denoted by $\mathbf{S}_E$.

Lemma. The Whitney test is equivalent to a particular case of the $\mathbf{S}_E$ test (3), if $h(z)$ is a linear decreasing function of the form $2h(z) \equiv C - z$, $z = 0, 1, \ldots, m\nu$, and to the test rejecting the hypothesis H_0 if $S_{Eq}^{+} - S_{Eq}^{-} > C - m\nu$, $q = 1, 2$, where C is the number involved into the Whitney test (1).

The proof directly follows from the equality (equivalence) of the events

$$\{S_{Eq}^{+} > (C - S_{Eq}^{0})/2\} = \{2S_{Eq}^{+} + S_{Eq}^{0} > C\} = \{U_q > C\}.$$

It is clear that some information about the samples can be lost as the pair of the statistics (S_{Eq}^{+}, S_{Eq}^{-}) is reduced to a simple difference $S_{Eq}^{+} - S_{Eq}^{-}$.

CHARACTERISTICS OF THE WHITNEY AND $\mathbf{S}_E$ TESTS

For calculating the level of significance of the Whitney test, we can use, for instance, Statement 1.

Let us use $\mathfrak{P}(u)$ to indicate the set of all ordered $(m+1)$ partitions $\mathfrak{p} = \mathfrak{p}(n) = (n_0, n_1, \ldots, n_m)$ of the number n (i.e., $n = n_0 + n_1 + \ldots + n_m$, where each integer number $n_i \geq 0$ and the sequence of the numbers n_i is essential) for which

$$\sum_{i=0}^{m}(m-i)n_i = u. \tag{4}$$

Let $\mathfrak{p}' = (n_0', n_1', \ldots, n_m') \in \mathfrak{P}(u_1)$ and $\mathfrak{p}'' = (n_0'', n_1'', \ldots, n_m'') \in \mathfrak{P}(u_2)$ be two $(m+1)$ partitions of the number n.

It is convenient to set $0! = 1$.

Statement 1 [6]. Let U_1 and U_2 be the statistics from Eq. (1). Then, if the hypothesis H_0 is valid, the probability is

$$\mathbf{P}\{U_1 = u_1, U_2 = u_2 \,|\, H_0\} = \frac{m!(n!)^2}{(m+2n)!} \sum_{\mathfrak{p}' \in \mathfrak{P}(u_1)} \sum_{\mathfrak{p}'' \in \mathfrak{P}(u_2)} \prod_{i=0}^{m} \binom{n_i' + n_i''}{n_i'}.$$

For the posed problem and in view of the fact that both the Whitney test and the proposed new test are most sensitive to shifts of distributions, it is of interest to compare them for families of distributions that differ mainly by shifts. Unfortunately, it is usually next to impossible to find explicit exact formulas for calculating the power and the MUF alarm probability of the tests. There are only some particular cases where they can be obtained. The simplest situations are the cases with exponential (this case is especially important for applications) and rectangular distributions of the form

$$H_0\colon F(x) = 1 - \mathrm{e}^{-x} \quad \text{for } x \geq 0, \tag{5}$$

$$H_1\colon G(x) = 1 - \mathrm{e}^{-(x-\theta)} \quad \text{for } x \geq \theta,\ \theta > 0 \tag{6}$$

and, correspondingly,

$$H_0\colon F(x) = x \quad \text{for } 0 \leq x \leq 1, \tag{7}$$

$$H_1\colon G(x) = \frac{x-\theta}{1-\theta} \quad \text{for } \theta \leq x \leq 1,\ 0 \leq \theta < 1. \tag{8}$$

For these distributions, we have $F(x_h) = b_0 + b_1 G(x_h)$,

$$\begin{aligned} 1 - F(x_h) &= b_1[1 - G(x_h)]; \\ F(x_h) - F(x_k) &= b_1[G(x_h) - G(x_k)], \quad x_h > x_k \geq \theta, \end{aligned} \tag{9}$$

where $b_0 = 1 - \exp(-\theta)$ for Eqs. (5) and (6) and $b_0 = \theta$ for Eqs. (7) and (8); in both cases, $b_1 = 1 - b_0$. This fact allows one to obtain explicit exact formulas both for the MUF alarm probability and for the power of the tests considered here.

The probability of the MUF alarm of the Whitney test, i.e., the probability of the hypothesis H_1 to be accepted in the case (let us denote it by H_1^*) where only some part of the neighborhood of the interfering object contour is actually present in the field of vision can be calculated with the use of the following statement.

Statement 2. In the case of H_1^*, let each of the variables Y_j have distribution (5) (or (7)) and let each of the random variables X_i and Z_j have distribution (6) (or (8)). Then the probability $\mathbf{P}\{U_1 = u_1, U_2 = u_2 \,|\, H_1^*\}$ with allowance for Eq. (9) can be presented as

$$m!(n!)^2 \sum_{\mathfrak{p}' \in \mathfrak{P}(u_1)} \sum_{\mathfrak{p}'' \in \mathfrak{P}(u_2)} \frac{1}{n_0'!n_0''!} \prod_{i=1}^{m} \binom{n_i' + n_i''}{n_i'} \sum_{r=0}^{n_0'} \binom{n_0'}{r} \frac{(n_0' + n_0'' - r)!}{(m+2n-r)!} b_0^r b_1^{n-r}. \tag{10}$$

Proof. As the functions F and G are assumed to be continuous, we can consider without loss of generality that the variables $X_1, \ldots, X_m$ have different values. Let us denote these variables arranged in increasing order by $X_{(1)} < X_{(2)} < \ldots < X_{(m)}$.

Let $x_1 < x_2 < \ldots < x_m$. We consider $m+1$ open intervals $I_0 = (-\infty, x_1)$, $I_h = (x_h, x_{h+1})$, $h = 1, 2, \ldots, m-1$, $I_m = (x_m, \infty)$. If there is an interfering object in the field of vision (at H_1^*), we obtain the following obvious equalities in accordance with the assumptions for the probability p_h that the random variable Y_j falls into the interval I_h, $h = 0, 1, \ldots, m$:

$$p_0 = F(x_1), \qquad p_h = F(x_{h+1}) - F(x_h), \quad h = 1,\, 2, \ldots, m-1,\ p_m = 1 - F(x_m);$$

for the probability q_h that the random variable Z_j falls into the interval I_h, $h = 0, 1, \ldots, m$, we have

$$q_0 = G(x_1), \qquad q_h = G(x_{h+1}) - G(x_h), \quad h = 1, 2, \ldots, m-1,\ q_m = 1 - G(x_m).$$

Thus, by virtue of the assumption of independence of observations, the probability of the events that exactly n_0' observations among all n observations Y_j and exactly n_0'' observations among all n observations Z_j are smaller than x_1, exactly n_1' and, respectively, n_1'' observations are located between x_1 and x_2, etc., and finally exactly n_m' observations among all n observations Y_j and exactly n_m'' observations among all n observations Z_j are greater than x_m, is equal to the product of two polynomial distributions

$$\frac{(n!)^2}{n_0'!n_1'!\ldots n_m'!n_0''!n_1''!\ldots n_m''!}\, p_0^{n_0'} p_1^{n_1'} \ldots p_m^{n_m'} q_0^{n_0''} q_1^{n_1''} \ldots q_m^{n_m''}.$$

If, in addition, we have $X_{(1)} = x_1, X_{(2)} = x_2, \ldots, X_{(m)} = x_m$, then Eq. (1) predicts that

$$U_1 = \sum_{i=0}^{m} (m-i) n_i', \quad U_2 = \sum_{i=0}^{m} (m-i) n_i''.$$

Therefore, we have

$$\mathbf{P}\{U_1 = u_1, U_2 = u_2 \,|\, X_{(1)} = x_1, X_{(2)} = x_2, \ldots, X_{(m)} = x_m; H_1^*\} =$$

$$= (n!)^2 \sum_{\mathfrak{p}' \in \mathfrak{P}(u_1)} \sum_{\mathfrak{p}'' \in \mathfrak{P}(u_2)} \Big(\prod_{k=0}^{m} n_h'! n_h''!\Big)^{-1} [F(x_1)]^{n_0'} [F(x_2) - F(x_1)]^{n_1'} \times$$

$$\times [F(x_3) - F(x_2)]^{n_2'} \ldots [F(x_m) - F(x_{m-1})]^{n_{m-1}'} [1 - F(x_m)]^{n_m'} \times$$

$$\times [G(x_1)]^{n_0''} [G(x_2) - G(x_1)]^{n_1''} \ldots [G(x_m) - G(x_{m-1})]^{n_{m-1}''} [1 - G(x_m)]^{n_m''}.$$

To obtain the probability $\mathbf{P}\{U_1 = u_1, U_2 = u_2 \,|\, H_1^*\}$, we have to integrate the last expression over the distribution of the ordered statistics $X_{(1)}, \ldots, X_{(m)}$ at H_1^*. The element of this distribution is $m!dG(x_1)\ldots dG(x_m)$ inside the domain, where $\theta < x_1 < x_2 < \ldots < x_m < \infty$ and is equal to zero outside the domain (see, e.g., [7]). From this equation, taking into account Eq. (9), we obtain ($n_i = n_i' + n_i''$, $i = 0, 1, \ldots, m$)

$$\mathbf{P}\{U_1 = u_1, U_2 = u_2 \,|\, H_1^*\} = m!(n!)^2 \sum_{\mathfrak{p}' \in \mathfrak{P}(u_1)} \sum_{\mathfrak{p}'' \in \mathfrak{P}(u_2)} \Big(\prod_{k=0}^{m} n_k'! n_k''!\Big)^{-1} b_1^{n - n_0'} \times$$

$$\times \int_{\theta}^{\infty} [1 - G(x_m)]^{n_m} dG(x_m) \int_{\theta}^{x_m} [G(x_m) - G(x_{m-1})]^{n_{m-1}} dG(x_{m-1}) \times$$

$$\times \int_{\theta}^{x_{m-1}} [G(x_{m-1}) - G(x_{m-2})]^{n_{m-2}} dG(x_{m-2}) \ldots \int_{\theta}^{x_3} [G(x_3) - G(x_{m-2})]^{n_2} dG(x_2) \times$$

$$\times \int_{\theta}^{x_2} [G(x_2) - G(x_1)]^{n_1} [b_0 + b_1 G(x_1)]^{n_0'} [G(x_1)]^{n_0''} dG(x_1).$$

Using now the Newtonian binomial formula, passing consecutively to new variables $y_1, \ldots, y_m$, and assuming that $y_i = G(x_i)$, $i = 1, \ldots, m$, we obtain

$$\mathbf{P}\{U_1 = u_1, U_2 = u_2 \,|\, H_1^*\} = m!(n!)^2 \sum_{\mathfrak{p}' \in \mathfrak{P}(u_1)} \sum_{\mathfrak{p}'' \in \mathfrak{P}(u_2)} \Big(\prod_{k=0}^{m} n_k'! n_k''! \Big)^{-1} \sum_{r=0}^{n_0'} \binom{n_0'}{r} b_0^r b_1^{n-r} \times$$

$$\times \int_0^1 [1 - y_m]^{n_m} dy_m \int_0^{y_m} [y_m - y_{m-1}]^{n_{m-1}} dy_{m-1} \int_0^{y_{m-1}} [y_{m-1} - y_{m-2}]^{n_{m-2}} dy_{m-2} \times \ldots$$

$$\ldots \times \int_0^{y_3} [y_3 - y_2]^{n_2} dy_2 \int_0^{y_2} [y_2 - y_1]^{n_1} y_1^{n_0 - r} dy_1. \quad (11)$$

According to Eqs. (3.7) and (3.9) from [1], the multiple integral in Eq. (11) is

$$\frac{(n_0 - r)! n_1! \ldots n_m!}{(n_0 - r + n_1 + n_2 + \ldots + n_m + m)!} = \frac{(n_0 - r)! n_1! \ldots n_m!}{(m + 2n - r)!}. \quad (12)$$

Substituting Eq. (12) into Eq. (11), we obtain presentation (10). Statement 2 is proved.

The probability of the MUF alarm of the Whitney test is

$$\sum_{u_1 > C} \sum_{u_2 > C} \mathbf{P}\{U_1 = u_1, U_2 = u_2 \,|\, H_1^*\}.$$

For calculating the power of the Whitney test, i.e., the probability of rejection of the basic hypothesis H_0 in the case where H_1 is valid (it is the important object that is present in the field of vision), we can use the following statement.

Statement 3 [3]. In the case of H_1, let each of the variables Y_j and Z_j have distribution (5) (or (7)), and let each of the random variables X_i have distribution (6) (or (8)). Then the probability $\mathbf{P}\{U_1 = u, U_2 = u_2 \,|\, H_1\}$ can be written in the following form in view of Eq. (9):

$$m!(n!)^2 \sum_{\mathfrak{p}' \in \mathfrak{P}(u_1)} \sum_{\mathfrak{p}'' \in \mathfrak{P}(u_2)} \prod_{i=0}^{m} \binom{n_i' + n_i''}{n_i'} \sum_{r=0}^{n_0' + n_0''} \frac{b_0^r b_1^{2n-r}}{(m + 2n - r)! r!}.$$

Let us now pass to the characteristics of the new $\mathbf{S}_E$ test (3). Let us first introduce necessary notations.

Let $\mathfrak{P}$ be a set of those ordered $(m+1)^2$ partitions $\mathfrak{p} = \mathfrak{p}(\nu)$ of the number ν of the form

$$\mathfrak{p}(\nu) = (\nu_{00}, \nu_{01}, \ldots, \nu_{0m}, \nu_{10}, \nu_{11}, \ldots, \nu_{1m}, \ldots, \nu_{m(m-1)}, \nu_{mm}),$$

for which the inequality $u > h(m\nu - u - v)$ is valid, where

$$u = \sum_{h=0}^{m-1} \sum_{k=0}^{m-1} (m - \max(h,k)) \nu_{hk}; \qquad v = \sum_{h=1}^{m} \sum_{k=1}^{m} \min(h,k) \nu_{hk}. \quad (13)$$

Let

$$\mathfrak{p}'(\nu) = (\nu'_{00}, \nu'_{01}, \ldots, \nu'_{0m}, \nu'_{10}, \nu'_{11}, \ldots, \nu'_{1m}, \ldots, \nu_{m(m-1)}, \nu'_{mm}) \in \mathfrak{P},$$

$$\mathfrak{p}''(\nu) = (\nu''_{00}, \nu''_{01}, \ldots, \nu''_{0m}, \nu''_{10}, \nu''_{11}, \ldots, \nu''_{1m}, \ldots, \nu''_{m(m-1)}, \nu''_{mm}) \in \mathfrak{P}$$

be two $(m+1)^2$ partitions of the number ν of this kind.

Statement 4 [3]. The level of significance of the $\mathbf{S}_E$ test is given by the formula

$$\mathbf{P}\{S^+_{Eq} > h(S^0_{Eq}),\ q = 1, 2 \,|\, H_0\} = \frac{m!(\nu!)^2}{(m+2n)!} \sum_{\mathfrak{p}' \in \mathfrak{P}} \sum_{\mathfrak{p}'' \in \mathfrak{P}} \Big(\prod_{k=0}^{m} s_k!\Big)\Big(\prod_{h,k=0}^{m} \nu'_{hk}!\nu''_{hk}!\Big)^{-1}, \tag{14}$$

where

$$s_k = \sum_{h=0}^{m} (\nu'_{kh} + \nu'_{hk} + \nu''_{kh} + \nu''_{hk}). \tag{15}$$

Let us consider the probability of the MUF alarm of the $\mathbf{S}_E$ test.

Statement 5. Let conditions of Statement 2 be satisfied. Then, in view of Eq. (9), the following inequality is valid:

$$\mathbf{P}\{S^+_q > h(S^0_q),\ q = 1, 2 \,|\, H^*_1\} =$$

$$= m!(\nu!)^2 \sum_{\mathfrak{p}' \in \mathfrak{P}} \sum_{\mathfrak{p}'' \in \mathfrak{P}} \Big(\prod_{k=1}^{m} s_k!\Big)\Big(\prod_{h,k=0}^{m} \nu'_{hk}!\nu''_{hk}!\Big)^{-1} \sum_{r=0}^{s'_0} \binom{s'_0}{r} \frac{(s_0 - r)! b_0^r b_1^{n-r}}{(m+2n-r)!},$$

where s_k, $k = 0, 1, \ldots, m$, as previously, is determined by formula (15):

$$s'_0 = \sum_{h=0}^{m} (\nu'_{0h} + \nu'_{h0}).$$

Proof. As previously, let $x_1 < x_2 < \ldots < x_m$ and, together with $m+1$ open intervals $I_0 = (-\infty, x_1)$, $I_k = (x_k, x_{k+1})$, $k = 1, 2, \ldots, m-1$, $I_m = (x_m, \infty)$. Let is consider $(m+1)^2$ open rectangles of the form $I_{hk} = I_h \times I_k$, $h, k = 0, 1, \ldots, m$.

By virtue of the assumptions made above (see also the beginning of the proof of Statement 2), if the hypothesis H^*_1 is valid, the probability that the pair $(Y_j, Y_{\nu+j})$ falls into the rectangle I_{hk} is $p_{hk} = p_h p_k$, while the probability that the pair $(Z_j, Z_{\nu+j})$ falls into the rectangle I_{hk} is $q_{hk} = q_h q_k$. Therefore, the probability that exactly ν'_{00} pairs among all ν pairs $(Y_j, Y_{\nu+j})$ and exactly ν''_{00} pairs among all ν pairs $(Z_j, Z_{\nu+j})$ fall into the rectangle I_{00}, and, respectively, exactly ν'_{hk} and ν''_{hk} pairs fall into the rectangle I_{hk} $(h, k = 0, 1, \ldots, m)$ is determined by the product of two polynomial distributions

$$\frac{\nu!}{\nu'_{00}!\nu'_{01}!\ldots\nu'_{0m}!\nu'_{10}!\nu'_{11}!\ldots\nu'_{m(m-1)}!\nu'_{mm}!} p_{00}^{\nu'_{00}} p_{01}^{\nu'_{01}} \ldots p_{0m}^{\nu'_{0m}} p_{10}^{\nu'_{10}} p_{11}^{\nu'_{11}} \ldots p_{m(m-1)}^{\nu'_{m(m-1)}} p_{mm}^{\nu'_{mm}} \times$$

$$\times \frac{\nu!}{\nu''_{00}!\nu''_{01}!\ldots\nu''_{0m}!\nu''_{10}!\nu''_{11}!\ldots\nu''_{m(m-1)}!\nu''_{mm}!} q_{00}^{\nu''_{00}} q_{01}^{\nu''_{01}} \ldots q_{0m}^{\nu''_{0m}} q_{10}^{\nu''_{10}} q_{11}^{\nu''_{11}} \ldots q_{m(m-1)}^{\nu''_{m(m-1)}} q_{mm}^{\nu''_{mm}} =$$

$$= (\nu!)^2 \Big(\prod_{h,k=0}^{m} \nu'_{hk}!\nu''_{hk}!\Big)^{-1} p_0^{s'_0} p_1^{s'_1} \ldots p_m^{s'_m} q_0^{s''_0} q_1^{s''_1} \ldots q_m^{s''_m},$$

where

$$s'_k = \sum_{h=0}^{m} (\nu'_{hk} + \nu'_{kh}), \qquad s''_k = \sum_{h=0}^{m} (\nu''_{hk} + \nu''_{kh}), \quad k = 0, 1, \dots, m.$$

From here, taking into account Eq. (9), we obtain ($s_i = s'_i + s''_i$, $i = 0, 1, \dots, m$)

$$\mathbf{P}\{S_q^+ > h(S_q^0),\ q = 1, 2 \,|\, X_{(1)} = x_1, X_{(2)} = x_2, \dots, X_{(m)} = x_m;\ H_1^*\} =$$

$$= (\nu!)^2 \sum_{\mathfrak{p}' \in \mathfrak{P}} \sum_{\mathfrak{p}'' \in \mathfrak{P}} \Big(\prod_{h,k=0}^{m} \nu'_{hk}!\nu''_{hk}! \Big)^{-1} [F(x_1)]^{s'_0} [F(x_2) - F(x_1)]^{s'_1} \times \dots$$

$$\dots \times [F(x_m) - F(x_{m-1})]^{s'_{m-1}} [1 - F(x_m)]^{s'_m} [G(x_1)]^{s''_0} [G(x_2) - G(x_1)]^{s''_1} \times \dots$$

$$\dots \times [G(x_m) - G(x_{m-1})]^{s''_{m-1}} [1 - G(x_m)]^{s''_m} =$$

$$= (\nu!)^2 \sum_{\mathfrak{p}' \in \mathfrak{P}} \sum_{\mathfrak{p}'' \in \mathfrak{P}} \Big(\prod_{h,k=0}^{m} \nu'_{hk}!\nu''_{hk}! \Big)^{-1} b_1^{n-s'_0} [b_0 + b_1 G(x_1)]^{s'_0} [G(x_1)]^{s''_0} \times$$

$$\times [G(x_2) - G(x_1)]^{s_1} \dots [G(x_m) - G(x_{m-1})]^{s_{m-1}} [(1 - G(x_m))]^{s_m}. \tag{16}$$

To obtain the probability $\mathbf{P}\{S_q^+ > h(S_q^0),\ q = 1, 2 \,|\, H_1^*\}$, we have to integrate Eq. (16) over the distribution of the variables $X_{(1)}, \dots, X_{(m)}$, whose element at H_1^*, as previously, is equal to $m!dG(x_1) \dots dG(x_m)$ inside the domain, where $\theta < x_1 < x_2 < \dots < x_m < \infty$, and to zero outside the domain. Let us first transform the expression under the sign of summation in Eq. (16) by omitting the product and using the Newtonian binomial formula to the expression

$$\sum_{r=0}^{s'_0} \binom{s'_0}{r} b_0^r b_1^{n-r} [G(x_1)]^{s_0 - r} [G(x_2) - G(x_1)]^{s_1} \times$$

$$\times [G(x_3) - G(x_2)]^{s_2} \dots [G(x_m) - G(x_{m-1})]^{s_{m-1}} [(1 - G(x_m))]^{s_m}. \tag{17}$$

Integrating now Eq. (17) over the distribution of the variables $X_{(1)}, \dots, X_{(m)}$ at H_1^*, consecutively passing to new variables $y_1, \dots, y_m$ by the formula $y_i = G(x_i)$, $i = 1, \dots, m$, and also using Eq. (11), the left side of Eq. (12), and the equalities

$$\sum_{k=0}^{m} s_k = \sum_{k=0}^{m} \sum_{h=0}^{m} (\nu'_{hk} + \nu'_{kh} + \nu''_{hk} + \nu''_{kh}) = 4\nu = 2n,$$

we obtain

$$m! \sum_{r=0}^{s'_0} \binom{s'_0}{r} b_0^r b_1^{n-r} \int_0^1 [1 - y_m]^{s_m} dy_m \int_0^{y_m} [y_m - y_{m-1}]^{s_{m-1}} y_{m-1} \times \dots$$

$$\dots \times \int_0^{y_3} [y_3 - y_2]^{s_2} dy_2 \int_0^{y_2} [y_2 - y_1]^{s_1} y_1^{s_0 - r} dy_1 = m! \Big(\prod_{k=1}^{m} s_k! \Big) \sum_{r=0}^{s'_0} \binom{s'_0}{r} \frac{(s_0 - r)! b_0^r b_1^{n-r}}{(m + 2n - r)!}.$$

Substituting this result into Eq. (16), we obtain the required result. Statement 5 is proved.

Let us now give the statement concerning the power of the $\mathbf{S}_E$ test.

Statement 6 [3]. Let the conditions of statement 3 be satisfied. Then, with allowance for Eq. (9), we have

$$\mathbf{P}\{S_q^+ > h(S_q^0),\ q = 1, 2 \,|\, H_1\} =$$

$$= m!(\nu!)^2 \sum_{\mathfrak{p}'(\nu) \in \mathfrak{P}} \sum_{\mathfrak{p}''(\nu) \in \mathfrak{P}} \Big(\prod_{k=1}^{m} s_k!\Big)\Big(\prod_{h,k=0}^{m} \nu'_{hk}!\nu''_{hk}!\Big)^{-1} \sum_{r=0}^{s_0} \binom{s_0}{r} \frac{(s_0 - r)! b_0^r b_1^{n-r}}{(m + 2n - r)!},$$

where s_k, $k = 0, 1, \ldots, m$, as previously, are determined by Eq. (15).

COMPARISON OF THE TEST CHARACTERISTICS. NUMERICAL EXAMPLES

Let us describe some numerical results that characterize the considered tests from different aspects to a certain extent. In [2], as suitable critical values of $h(z)$, $z = 0, \ldots, m\nu$, of the $S_{Eq}^+ > h(S_{Eq}^0)$ test used as a basis for constructing the $\mathbf{S}_E$ test (3), we proposed to use the critical values obtained with the use of the known Neumann–Pearson lemma for the case with an alternative close to the zero hypothesis of the form

$$G = (1 - a)F + aF^2. \tag{18}$$

In choosing Eq. (18), we took into account the fact that, with this alternative, for sufficiently small values of $a > 0$, the MW test $U_q > C$, being equivalent to the Wilcoxon test, is the most powerful (ensures the highest probability that the hypothesis H_1 is accepted when it is valid) among the rank tests (see, e.g., [8, Chapter 6, Section 12, Problem 28]). For the critical values of $h(z)$, $z = 0, \ldots, m\nu$, obtained for a number of fixed values a^* of the parameter a, the test $S_{Eq}^+ > h(S_{Eq}^0)$ turned out to be more powerful that the MW test in a fairly wide range of the values of the parameter a in Eq. (18), namely, for $a > \gamma(m, n)$, where the boundary value of γ decreases with increasing m and n (see [2] for more details).

Case with $m = 5$ and $n = 2\nu = 4$

Let $C = 15$ in Eq. (1). At this value of C, the level of significance of the MW test is equal to 0.0(952380), which is an infinite periodic fraction, and the level of significance of the Wh (1) turned out to be 0.0(257904). In [2], suitable critical values of $h(z)$, $z = 0, 1, \ldots, 10$, of the S_E test turned out to be criteria obtained, in particular, at $a^* = 0.53$. They are listed in Table 1. At these values, the level of significance of the S_E test is 0.0(925), and the level of significance of the $\mathbf{S}_E$ test is 0.0(264698), i.e., it is slightly higher than the level of significance of the Wh test. Therefore, it was only for a more correct comparison of the powers of the tests that we made the replacement $h(0) = 9$, after which the level of significance of the $\mathbf{S}_E$ test (3) became equal to 0.0(250268), which is lower than the level of significance of the Wh test.

The calculated powers and MUF alarm probabilities of the Wh- and $\mathbf{S}_E$ tests for different values of θ in distributions (6) and (8) are listed in Tables 2 and 3. The values in these tables are given with sufficient accuracy for comparisons; the first and second rows refer to the Wh test, while the third and fourth rows refer to the $\mathbf{S}_E$ test. The first and third rows show the MUF alarm probabilities of the tests, while the test power values are given in the second and fourth rows. The column with $\theta = 0.0$ gives the levels of significance of the tests.

Table 1. Values of $h(z)$, $z = 0, 1, \ldots, 10$, of the $\boldsymbol{S}_E$ test at $m = 5$, $n = 4$

z	0	1	2	3	4	5	6	7	8	9	10
h	8	7	7	6	5	4	4	3	2	1	0

Table 2. Powers and MUF alarm probabilities of the Wh and $\mathbf{S}_E$ tests in the case with $m = 5$, $n = 4$, and alternative (6)

θ										
0.0	0.01	0.1	0.2	0.5	1.0	1.5	2.0	2.5	3.0	4.0
0.0257	0.02639	0.0317	0.0377	0.0539	0.072	0.0832	0.088	0.091	0.093	0.0945
0.0257	0.02701	0.0404	0.0614	0.1660	0.419	0.6440	0.791	0.878	0.928	0.9749
0.0250	0.02565	0.0313	0.0377	0.0550	0.074	0.0834	0.087	0.088	0.089	0.0898
0.0250	0.02630	0.0410	0.0657	0.1995	0.523	0.7737	0.904	0.962	0.985	0.9979

Table 3. Powers and MUF alarm probabilities of the Wh and $\mathbf{S}_E$ tests in the case with $m = 5$, $n = 4$, and alternative (8)

θ									
0.0	0.01	0.1	0.2	0.3	0.5	0.7	0.8	0.9	0.95
0.0257	0.02639	0.0321	0.0391	0.046	0.062	0.077	0.0847	0.090	0.093
0.0257	0.02701	0.0413	0.0672	0.108	0.259	0.519	0.6827	0.849	0.928
0.0250	0.02565	0.0317	0.0392	0.047	0.063	0.078	0.0846	0.088	0.089
0.0250	0.02631	0.0421	0.0720	0.124	0.321	0.640	0.8113	0.945	0.985

Table 4. Values $h(z)$, $z = 0, 1, \ldots, 21$, of the $\mathbf{S}_E$ test at $m = 7$ and $n = 6$

z	0	1	2	3	4	5	6	7	8	9	10
h	15	15	14	14	13	13	12	11	11	10	9
z	11	12	13	14	15	16	17	18	19	20	21
h	9	8	7	6	6	5	4	3	2	1	0

Table 5. Powers and MUF alarm probabilities of the Wh and $\mathbf{S}_E$ tests in the case with $m = 7$, $n = 6$, and alternative (6)

θ									
0.0	0.01	0.1	0.2	0.5	1.0	1.5	2.0	2.5	3.0
0.0249	0.0257	0.0327	0.0403	0.0600	0.0789	0.086	0.088	0.0898	0.0901
0.0249	0.0265	0.0447	0.0750	0.2355	0.5918	0.828	0.935	0.9763	0.9914
0.0244	0.0252	0.0325	0.0404	0.0604	0.0783	0.084	0.086	0.0874	0.0876
0.0244	0.0260	0.0455	0.0793	0.2618	0.6433	0.867	0.955	0.9852	0.9950

Case with $m = 7$ and $n = 2\nu = 6$

At $C = 30$, the level of significance of the MW test is equal to 0.0903 (with accuracy to four decimal digits), and the accuracy of the Wh test is 0.0249. Similar to the previous case, the critical conditions of the S_E test were first obtained at $a^* = 0.55$; these critical values are listed in Table 4, and the level of significance of this test obtained with these values turned out to be 0.0898.

For the level of significance of the $\mathbf{S}_E$ test to be lower than the level of significance of the Wh test, we applied the replacement $h(14) = 7$.

The powers and the MUF alarm probabilities for different values of θ in Eqs. (6) and (8) in this case are compared in Tables 5 and 6. As previously, the first and second rows refer to the Wh test, while the third and fourth rows refer to the $\mathbf{S}_E$ test. The first and third rows give the values of the MUF alarm probabilities of the tests, while the test powers are indicated in the second and fourth rows.

Comparing the values listed in the tables, we can clearly see that, as θ departs from $\theta_0 = 0$ in Eqs. (6) and (8), the power and the MUF alarm probability of the $\mathbf{S}_E$ test rapidly increase and become greater than the corresponding values of the Wh test, though it should be noted that only slightly greater in the case of the MUF alarm probability. However, this occurs only at moderately high values of θ, where the values of the power and MUF alarm probability are so small that they are not of interest for practice. The observers (users) usually are not interested in the test power in the cases of alternatives with θ close to θ_0; instead, they want to control the probability of detection of alternatives that are rather far away from θ_0 [8, p. 145].

Table 6. Powers and MUF alarm probabilities of the Wh and $\mathbf{S}_E$ tests in the case with $m = 7$, $n = 6$, and alternative (8)

θ									
0.0	0.01	0.1	0.2	0.3	0.5	0.6	0.8	0.9	0.95
0.0249	0.02575	0.0331	0.0420	0.0514	0.0693	0.0768	0.087	0.0896	0.0901
0.0249	0.02656	0.0460	0.0837	0.1467	0.3752	0.5366	0.860	0.9647	0.9913
0.0244	0.02522	0.0329	0.0392	0.0517	0.0694	0.0764	0.085	0.0872	0.0876
0.0244	0.02607	0.0470	0.0892	0.1610	0.4164	0.5872	0.895	0.9771	0.9949

As θ moves further away from θ_0, the growth rate of the MUF alarm probability of the $\mathbf{S}_E$ test noticeably decreases, as is seen from Tables 2, 3, 5, and 6. As a result, this probability at $\theta > \theta^*(m, n)$ becomes lower than the MUF alarm probability of the Wh test. It should be also noted that the boundary value $\theta^* = \theta^*(m, n)$ decreases with increasing m and n.

CONCLUSION

In this paper, we consider a new nonparametric statistical test for problems with three samples. A comparison of the new test with the Whitney test in the cases with exponential and uniform alternative distributions shows that the power (in the detection problem, the detection probability) of the new test is noticeably higher than the power of the Whitney test. The probability of the most undesirable false alarm, where some part of the neighborhood of the contour of an interfering object of a greater size is observed in the field of vision of the observed instead of the specified important object, did not turn out to be appreciably higher than a similar probability predicted by the Whitney test; vice versa, the probability was even somewhat lower than the latter (which is an unexpected wonderful fact). As nonparametric tests are comparatively little sensitive to deviations from particular alternative distributions used for constructing the tests, we can hope that these two advantages of the new test will be also manifested in many other cases, not only in the cases with exponential and uniform distributions analyzed here. This is extremely important for many applications under versatile conditions.

This work was supported by the Presidium of the Russian Academy of Sciences (Program No.32) and by the Russian Foundation for Basic Research (Grant No. 13-07-00068).

REFERENCES

1. G. I. Salov, "New Statistical Test for Problems with Two and Three Samples, which is More Powerful than the Wilcoxon and Whitney Tests," Avtometriya **47** (4), 58–70 (2011) [Optoelectron., Instrum. Data Process. **47** (4), 368–377 (2011)].
2. G. I. Salov, "On the Power of a New Statistical Test and Two-Sample Wilcoxon Test," Avtometriya **50** (1), 44–59 (2014) [Optoelectron., Instrum. Data Process. **50** (1), 36–48 (2014)].
3. G. I. Salov, "New Nonparametric Statistical Test for Problems with Three Samples whose Particular Case is Equivalent to the Whitney Test," Sib. Zh. Vych. Mat. **17** (4), 389–397 (2014).
4. D. R. Whitney, "A Bivariate Extension of the U Statistic," Ann. Math. Stat. **22** (2), 274–282 (1951).
5. H. B. Mann and D. R. Whitney, "On a Test of whether One of Two Random Variables is Stochastically Larger than the Other," Ann. Math. Stat. **18** (1), 50–60 (1947).
6. G. I. Salov, "On the Power of Nonparametric Tests for Detection of Extended Objects on a Random Background," Avtometriya, No. 3, 60–75 (1997).
7. S. S. Wilks, "Order Statistics," Bull. Amer. Math. Soc. **54** (1), 5–50 (1948).
8. E. Lehman, *Testing Statistical Hypotheses*, Wiley (1959).