ON A RESULT BY M. ROSENBLATT CONCERNING THE VON MISES-SMIRNOV TEST

By M. Fisz

University of Warsaw

1. Summary. Rosenblatt's derivation [3] of the limiting distribution of the statistic (1) below contains an incorrect step.[1] A simple argument is presented that corrects Rosenblatt's proof, so that his conclusion is shown to be valid.

2. Rosenblatt's result. Let x_k $(k = 1, \cdots, n)$ and y_j $(j = 1, \cdots, m)$ be two independent random samples from two populations with the same continuous distribution function $F(t)$. Let $S_1(t)$ and $S_2(t)$ denote the corresponding empirical distribution functions. Lehmann [2] has suggested

$$(1) \qquad (mn/(n+m)) \int_{-\infty}^{\infty} [S_1(t) - S_2(t)]^2 \, d[(nS_1(t) + mS_2(t))/(n+m)]$$

as a test statistic for the two sample problem. Rosenblatt [3] has proved that the statistic (1) has the same limiting distribution, when $n \to \infty$, $m \to \infty$, $m/n \to \lambda > 0$, as the von Mises-Smirnov statistic (Smirnov [4]),

$$n \int_{-\infty}^{\infty} [S(t) - F(t)]^2 \, dF(t).$$

An essential role in Rosenblatt's proof is played by the equality

$$(2) \qquad \begin{aligned} (nm/(n+m)) \int_0^1 [S_1(t) - S_2(t)]^2 \, d[(nS_1(t) + mS_2(t))/(n+m) - t] \\ = (nm/(n+m)) \left\{ \int_0^1 [S_2(t) - t]^2 \, d[S_1(t) - t] \right. \\ \left. + \int_0^1 [S_1(t) - t]^2 \, d[S_2(t) - t] \right\}, \end{aligned}$$

where the non-restrictive assumption has been made that $F(t)$ is the uniform distribution function on [0, 1]. Now simple calculations show that (2) does not hold. Set

$$(3) \qquad A = \int_0^1 [S_1(t) - S_2(t)]^2 \, d[(nS_1(t) + mS_2(t))/(n+m)],$$

$$(4) \qquad B = \int_0^1 [S_1(t) - S_2(t)]^2 \, dt,$$

$$(5) \qquad C = \int_0^1 [S_2(t) - t]^2 \, d[S_1(t) - t] + \int_0^1 [S_1(t) - t]^2 \, d[S_2(t) - t].$$

Received April 17, 1959; revised February 18, 1960.

[1] This has been noted in a paper by J. Kiefer [1], which appeared after the present note was submitted.

Let us assume that $S_1(t)$ and $S_2(t)$ are continuous from the right. We have then (with probability 1, since Pr $(x_{k_1} \neq x_{k_2} \neq y_{j_1} \neq y_{j_2}, k_1, k_2 = 1, \cdots, n, j_1, j_2 = 1, \cdots, m, k_1 \neq k_2, j_1 \neq j_2) = 1)$ that,

$$\begin{aligned} A = [1/(n+m)]\Big\{\sum_{k=1}^{n}[(k/n)^2 - 2(k/n)S_2(x_k) + S_2^2(x_k)] \\ + \sum_{j=1}^{m}[(j/m)^2 - 2(j/m)S_1(y_j) + S_1^2(y_j)]\Big\}, \end{aligned} \tag{6}$$

$$\begin{aligned} C &= (1/n)\sum_{k=1}^{n}[S_2(x_k) - x_k]^2 + (1/m)\sum_{j=1}^{m}[S_1(y_j) - y_j]^2 \\ &\quad - \int_0^{x_1} t^2\,dt - \sum_{k=1}^{n-1}\int_{x_k}^{x_{k+1}}[(k/n) - t]^2\,dt - \int_{x_n}^{1}(1-t)^2\,dt \\ &\quad - \int_0^{y_1} t^2\,dt - \sum_{j=1}^{m-1}\int_{y_j}^{y_{j+1}}[(j/m) - t]^2\,dt - \int_{y_m}^{1}(1-t)^2\,dt \\ &= (1/n)\sum_{k=1}^{n}S_2^2(x_k) + (1/m)\sum_{j=1}^{m}S_1^2(y_j) - \tfrac{2}{3} \\ &\quad + (1/n)^2\sum_{k=1}^{n}[2k - 1 - 2nS_2(x_k)]x_k \\ &\quad + (1/m)^2\sum_{j=1}^{m}[2j - 1 - 2mS_1(y_j)]y_j. \end{aligned} \tag{7}$$

We find from (6) that

$$\begin{aligned} A &- (1/n)\sum_{k=1}^{n}S_2^2(x_k) - (1/m)\sum_{j=1}^{m}S_1^2(y_j) + \tfrac{2}{3} \\ &= [1/(n+m)]\Big\{\sum_{k=1}^{n}[(k/n)^2 - 2(k/n)S_2(x_k) - (m/n)S_2^2(x_k)] \\ &\quad + \sum_{j=1}^{m}[(j/m)^2 - 2(j/m)S_1(y_j) - (n/m)S_1^2(y_j)]\Big\} + \tfrac{2}{3} \\ &= [1/(n+m)]\Big[\sum_{k=1}^{n}(k/n)^2 + \sum_{j=1}^{m}(j/m)^2 + (1/(nm))\Big(\sum_{k=1}^{n}k^2 + \sum_{j=1}^{m}j^2\Big)\Big] \\ &\quad - \frac{n+m}{nm}\Big\{\sum_{k=1}^{n}\Big[\frac{nS_1(x_k) + mS_2(x_k)}{n+m}\Big]^2 + \sum_{j=1}^{m}\Big[\frac{nS_1(y_j) + mS_2(y_j)}{n+m}\Big]^2\Big\} + \tfrac{2}{3} \\ &= [1/(n+m)]\Big[\sum_{k=1}^{n}(k/n)^2 + \sum_{j=1}^{m}(j/m)^2 + (1/(nm))\Big(\sum_{k=1}^{n}k^2 + \sum_{j=1}^{m}j^2\Big)\Big] \\ &\quad - [(n+m)/nm]\sum_{r=1}^{n+m}[r/(n+m)]^2 + \tfrac{2}{3} = 1/(6nm). \end{aligned} \tag{8}$$

On the other hand, we have

$$
(9)\quad \begin{aligned} B &= \int_0^1 S_1^2(t)\,dt + \int_0^1 S_2^2(t)\,dt - 2\int_0^1 S_1(t)S_2(t)\,dt \\ &= -(1/n)^2\sum_{k=1}^{n}[2k-1-2nS_2(x_k)]x_k - (1/m)^2\sum_{j=1}^{m}[2j-1-2mS_1(y_j)]y_j . \end{aligned}
$$

Relations (7)–(9) imply that $A - B - C = 1/(6nm)$. Consequently the left side of (2) differs from the right one by $1/[6(n+m)]$.

Although equality (2) does not hold, the assertion of Rosenblatt's theorem remains true, since $1/[6(n+m)] \to 0$ as $n \to \infty$ and $m \to \infty$.

REFERENCES

[1] J. KIEFER, "K-sample analogues of the Kolmogorov-Smirnov and Cramér-v. Mises tests," *Ann. Math. Stat.*, Vol. 30 (1959), pp. 420–447.

[2] E. L. LEHMANN, "Consistency and unbiasedness of certain nonparametric tests," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 165–179.

[3] M. ROSENBLATT, "Limit theorems associated with variants of the von Mises statistic," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 617–623.

[4] N. SMIRNOFF, "Sur la distribution de ω^2," *Comptes Rendus de l'Academie des Sciences*, Vol. 202 (1936), pp. 449–452.