A Note on a Nonparametric Two-Sample Test

Ishii, Goro
Atomic Bomb Casuality Comission

https://doi.org/10.5109/12994

出版情報：統計数理研究. 9 (2/3), pp.57-60, 1960-10. Research Association of Statistical Sciences

A NOTE ON A NONPARAMETRIC TWO-SAMPLE TEST

By

Goro ISHII

(Received December 5, 1959)
(Revised January 31, 1960)

1. Introduction

Let X and Y be the two random variables which follow the distribution functions of continuous type $F(x)$ and $G(y)$, respectively. The test of the hypothesis H_0: $F(x)=G(x)$ is treated here. For this purpose a slightly modified test of David-Okamoto's test [2], [6] will be applied.

Let x_i, $i=1, 2, \cdots, N$, and y_j, $j=1, 2, \cdots, M$, be N and M observations of X and Y. These x_i and y_j are mixed and arranged in order of magnitude. Denote these by

(1) $$z_1 < z_2 < \cdots\cdots\cdots\cdots < z_L$$

where $L=N+M$. We shall consider here only the case when $L=n\alpha$ where α is a fixed integer. Divide (1) into n classes from the left so that each class contains α members. That is

(2) $$z_1, z_2, \cdots, z_\alpha ; z_{\alpha+1}, \cdots, z_{2\alpha} ; \cdots ; \cdots, z_{n\alpha}.$$

The number of x_i involved in each clase is denoted by i_k, $k=1, 2, \cdots, n$. Consider a function of i

(3) $$h(i)=0, \quad \text{when } i \neq 0,$$
$$=1, \quad \text{when } i=0.$$

Then

$$v=\sum_{k=1}^{n} h(i_k)$$

is the number of classes which contain no x's. And $u=n-v$ is the number of classes which contain some x's.

A large value of v suggests that the hypothesis should be rejected.

Therefore we use the v as a test function and call this test v-test.

In many occations, a test of goodness of fit can be extended to a two sample test, Darling [1]. For example, Kolmogorov's test of fit by $D_n=\sup|F(x)-F_n(x)|$ has a corresponding two sample test by $D_{n,m}=\sup|F_n(x)-G_m(x)|$ while ω^2 test of fit by $\omega_n^2=n\int_{-\infty}^{\infty}(F(x)-F_n(x))^2dF(x)$ has a corresponding two sample test by $\omega^2_{n,m}=\frac{mn}{m+n}\int_{-\infty}^{\infty}(F_n(x)-G_m(x))^2 d\frac{nF_n+mG_m}{n+m}$.

David [2] proposed a test of fit by the following procedure. Let x_i, $i=1,2,\cdots,N$, be N independent observations of X. Since $F(x)$ is continuous, there are real numbers $\{a_i\}$, $i=0,1,2,\cdots,n-1$ such that $F(a_{i+1})-F(a_i)=1/n$ where $a_0=-\infty$ and $a_n=+\infty$. Let v_0 be the number of intervals (a_{i-1}, a_i) which contain no x's. If v_0 is too large, the hypothesis that the distribution of X is $F(x)$ will be rejected. The detailed character of this test was treated by Okamoto [6], Weiss [7], Kitabatake [5], and the application of the test to life testing was treated by Ishii [4]. It is clear that the v-test above mentioned is a two sample test corresponding to the test of fit by David-Okamoto.

2. Distribution of v

Let us consider the behavior of x's in series (2). The probability that Nx's are contained in at most k classes is

$$\binom{k\alpha}{N}\Big/\binom{L}{N}.$$

This expression has a meaning only when $k\alpha\geqq N$. When $k\alpha<N$, we let the probability 0.

Denote by p_i the probability that i classes in the above k classes contain all of the Nx's and that each contains at least one of the x's. Then we have

$$\binom{k\alpha}{N}\Big/\binom{L}{N}=\sum_{i=1}^{k}\binom{k}{i}p_i.$$

Therefore for every positive integer ν

$$\sum_{k=1}^{\nu}(-1)^{\nu-k}\binom{\nu}{k}\binom{k\alpha}{N}\Big/\binom{L}{N}$$
$$=\sum_{k=1}^{\nu}(-1)^{\nu-k}\binom{\nu}{k}\sum_{i=1}^{k}\binom{k}{i}p_i$$
$$=\sum_{i=1}^{\nu}p_i\binom{\nu}{i}\sum_{k=i}^{\nu}(-1)^{\nu-k}\binom{\nu-i}{k-i}=p_\nu.$$

Thus

$$P\{u=\nu\}=\binom{n}{\nu}p_\nu=\binom{n}{\nu}\sum_{k=1}^{\nu}(-1)^{\nu-k}\binom{\nu}{k}\binom{k\alpha}{N}\Big/\binom{L}{N}.$$

Then

$$P\{v=\nu\}=\binom{n}{\nu}\sum_{k=1}^{n-\nu}(-1)^{n-\nu-k}\binom{n-\nu}{k}\binom{k\alpha}{N}\Big/\binom{L}{N}\qquad \nu=0,1,\cdots,n-1. \tag{4}$$

The s-th factorial moment of v is

$$E(v^{(s)})=\frac{((n-s)\alpha)^{(N)}\,n!}{(n\alpha)^{(N)}\,(n-s)!}, \tag{5}$$

where

$$x^{(s)}=x(x-1)\cdots(x-s+1).$$

Putting $s=1, 2$

$$(6) \qquad E(v)=n\frac{(L-\alpha)^{(N)}}{L^{(N)}}$$

$$E(v(v-1))=n(n-1)\frac{(L-2\alpha)^{(N)}}{L^{(N)}}.$$

In the case when N is fixed and M tends to infinity, the probability (4) tends to $p(v=\nu)$ in Okamoto [6].

In the case when both N and M tend to infinity, we can assume many conditions on the increasing order of N, M, α and n. In the following we shall treat the case that α and n tend to infinity with order $\sqrt{L}$ and $M=n(\alpha-r)$, $N=nr$ where r is fixed. In this case the factorial moment (5) becomes

$$E(v^{(s)})\sim\left(1-\frac{s}{n}\right)^{nr}\frac{n!}{(n-s)!}.$$

On the asymptotic behavior of v the following theorem is developed. The proof is similar to that of theorem 1 in Okamoto [6] and theorem 1 in Ishii [4]. Therefore, the proof is omitted.

Theorem 1.

v/n is asymptotically normally distributed with mean e^{-r} and variance $e^{-2r}(e^{r}-1-r)/n$, where $r=N/n$ =const.

3. Application to life test.

Suppose that life testing of X and Y start with Nx items and My items and stops when $L_1=n_1\alpha$ deaths in total occur, where L_1 is a certain preassigned number. Let z_i, $i=1, 2, \cdots, L_1$, be the ordered observations of X and Y. Denote by N_1 and M_1 the number of X and Y in $L_1 z$'s.

$$L_1=n_1\alpha=N_1+M_1$$
$$L=n\alpha=N+M$$
$$L_1\leqq L.$$

Divide these z_i into n_1 classes in the same way as in section 1

$$(7) \qquad z_1, z_2, \cdots, z_\alpha; z_{\alpha+1}, \cdots, z_{2\alpha}; \cdots, z_{n_1\alpha}.$$

Let v_1 be the number of classes that contain no x's in the above n_1 classes. We shall use the v_1 as a test function of testing the hypothesis H_0.

N_1 follows the hypergeometric distribution

$$p\{N_1\}=\binom{N}{N_1}\binom{M}{M_1}\Big/\binom{L}{L_1}.$$

The probability of v_1 under the condition that N_1 is fixed is

$$(8)\quad P\{v_1=\nu N_1\}=\frac{\binom{n_1}{\nu}}{\binom{L_1}{N_1}}\sum_{k=1}^{n_1-\nu}(-1)^{n_1-\nu-k}\binom{n_1-\nu}{k}\binom{k\gamma}{N_1}\,.$$

Then

$$(9)\quad P\{v_1=\nu, N_1\}=P\{v_1=\nu N_1\}\cdot P\{N_1\}$$

$$(10)\quad P\{v_1=\nu\}=\sum_{N_1}P\{v_1=\nu, N_1\}\,.$$

The s-th factorial moment of v_1 is

$$E(v_1^{(s)})=\frac{((n-s)\alpha)^{(N)}n_1!}{(n\alpha)^{(N)}(n_1-s)!}\,.$$

Putting $s=1,\ 2$

$$E(v_1)=n_1\frac{(L-\alpha)^{(N)}}{L^{(N)}}$$

$$E(v_1(v_1-1))=n_1(n_1-1)\frac{(L-2\alpha)^{(N)}}{L^{(N)}}\,.$$

In the case when N and n_1/n are fixed and M tends to infinity, the probability (9) tends to $P(v_1=\nu)$ in Ishii [4]. In the case when both N and M tend to infinity, we shall add a condition that $n_1/n=t$ is fixed to the assumptions in section 2. Under these assumptions we have the following theorem on the asymptotic behavior of v_1. The proof is also similar to that of theorem 1 in Ishii [4]. Therefore it is omitted.

Theorem 2.

v_1/n_1 is asymptotically normally distributed with mean e^{-r} and variance $e^{-2r}(e^r-1-tr)/n$, where $r=N/n$ and $t=n_1/n$ are constants.

Atomic Bomb Casualty Comission, Hiroshima

References

1. D. A. Darling, "*The Kolmogorov-Smirnov, Cramer-Von Mises Tests*" Ann. Math. Stat. Vol. **28** (1957), pp. 823-838.
2. F. N. David, "*Two combinatorial tests of whether a sample has come from a given population*" Biometrika Vol. **37** (1950), pp. 97-110.
3. G. Ishii, "*Test of fit in life test*" Ann. Inst. Stat. Math. Vol. **9**(1957-8), pp. 117-125.
4. G. Ishii, "*On a non-parametric test in life test*" Bull. Math. Stat. Vol. 8 (1959), pp. 73-79.
5. S. Kitabatake, "*A remark on a non-parametric test*" Mathematica Japonicae, Vol. **5** (1958), pp. 45-49.
6. M. Okamoto, "*On a non-parametric test*" Osaka Math. J. Vol. **4** (1952), pp 76-85.
7. I. Weiss, "*Limiting distribution in some occupancy problem*" Ann. Math. Stat. Vol. **29** (1958), pp. 878-884.