数理解析研究所講究録 1224 巻 2001 年 182-186

A simulation study on testing the hypothesis in the two-sample problem

Masafumi Akahira and Kunihiko Takahashi

赤平 昌文 (筑波大・数学)　　髙橋 邦彦 (筑波大・数学)

Institute of Mathematics, University of Tsukuba, Ibaraki 305–8571, Japan

Abstract

We consider the problem of comparing two population distributions F and G based on two samples $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$, one from each population. In testing the hypothesis $H : F \equiv G$, we assume that the size n of the sample is large, but m is small. Then we apply the Kolmogorov-Smirnov type test to the case using the resampling method including the bootstrap, and carry out a simulation study on the power of the test under some population distributions.

1. Kolmogorov-Smirnov type test

Suppose that $X_1, \ldots, X_n$ are independent and identically distributed (i.i.d.) random variables according to the cumulative distribution function (c.d.f.) F and $Y_1, \ldots, Y_m$ are i.i.d. random variables according to the c.d.f. G. In the case when n is large but m is small, we consider the problem of testing hypothesis $H : F \equiv G$. Now, let F_m be the empirical distribution function (e.d.f.) based on resampling $X_1', \ldots, X_m'$ with replacement from $X_1, \ldots, X_n$ and G_m be the e.d.f. based on the bootstrap sample $Y_1', \ldots, Y_{m-1}'$ of size $m-1$ from the e.d.f. F_m. Here, we take $m-1$ instead of m from the viewpoint of the unbiasedness. (For details, see Akahira and Takeuchi (1991), page 300). Then we consider the Kolmogorov-Smirnov type test of level α. For any α $(0 < \alpha < 1)$ there exists c such that

$$P\left\{\sup_x |F_m(x) - G_m(x)| \geq \frac{c}{m}\right\} = \alpha.$$

Indeed, for $\alpha = 0.05, 0.01$, $m = 1(1)15$, $20(5)30$, $40(10)100$, the values of c are given by Birnbaum (1952) (see also Miller (1956)).

Now, letting N be the repeated number of resampling $X'_1, \ldots, X'_m$ with replacement from $X_1, \ldots, X_n$, for the e.d.f. G_m we denote by k_N the frequency number satisfying

$$\sup_x |F_m(x) - G_m(x)| \geq c/m$$

under the hypothesis $H : F \equiv G$. Then we make a rule on testing hypothesis as follows. If k_N/N α, then one rejects the hypothesis H, otherwise one accepts it. In a similar way to the above, under the alternative hypothesis $K : F \not\equiv G$, we may regard k_N/N as the power of the test with the rejection region. For example, in practice the above way may be applied to the test of the effect of a drug.

2. Simulation study

In testing the hypothesis in the previous section we consider the following cases.
(i) F is the c.d.f. of the beta distribution $Be(a, b)$ with the probability density function (p.d.f.)

$$f(x) = \begin{cases} \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1} & \text{for } 0 < x < 1, \\ 0 & \text{otherwise,} \end{cases}$$

where $a > 0$ and $b > 0$. Let $\alpha = 0.05$. When G is the c.d.f. of the beta distribution $Be(4,4)$, for $(a,b) = (4,5), (4,4.5), (4,4.2)$ and $N = 500$, we have the values of the approximate power k_N/N as in Table 1. (See also Figure 1.) When G is the c.d.f. of the beta distributin $Be(3,3)$, for $(a,b) = (3,3.2), (3,3.5), (3,4), (3,5)$ and $N = 500$, we have the values of the approximate power k_N/N as in Table 2. (See also Figure 2.)

Table 1 *Comparison of the approximate power k_N/N of the test in three simulation results in case of $Be(4,4)$*

(a,b)	m	k_N/N		
(4, 5)	100	0.780	0.930	0.858
(4, 5)	50	0.982	0.880	0.772
(4, 5)	30	0.466	0.632	0.726
(4, 4.5)	100	0.224	0.762	0.756
(4, 4.5)	50	0.424	0.178	0.270
(4, 4.5)	30	0.116	0.220	0.306
(4, 4.2)	100	0.178	0.152	0.268
(4, 4.2)	50	0.178	0.232	0.102
(4, 4.2)	30	0.062	0.196	0.200

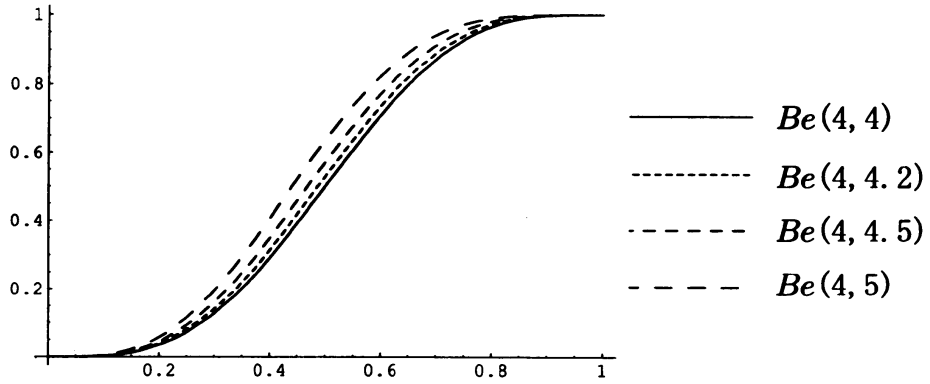

Figure 1 *The c.d.f.'s of the beta distributions*

Table 2 *Comparison of the approximate power k_N/N of the test in three simulation results in case of $Be(3,3)$*

(a,b)	m	k_N/N		
(3, 3.2)	100	0.430	0.110	0.194
(3, 3.2)	50	0.336	0.350	0.154
(3, 3.2)	30	0.088	0.494	0.700
(3, 3.5)	100	0.312	0.948	0.552
(3, 3.5)	50	0.522	0.372	0.596
(3, 3.5)	30	0.388	0.190	0.936
(3, 4)	100	0.808	0.988	0.898
(3, 4)	50	0.924	0.732	0.602
(3, 4)	30	0.184	0.734	0.902
(3, 5)	100	1.000	1.000	1.000
(3, 5)	50	0.972	0.994	0.986
(3, 5)	30	0.998	0.998	0.838

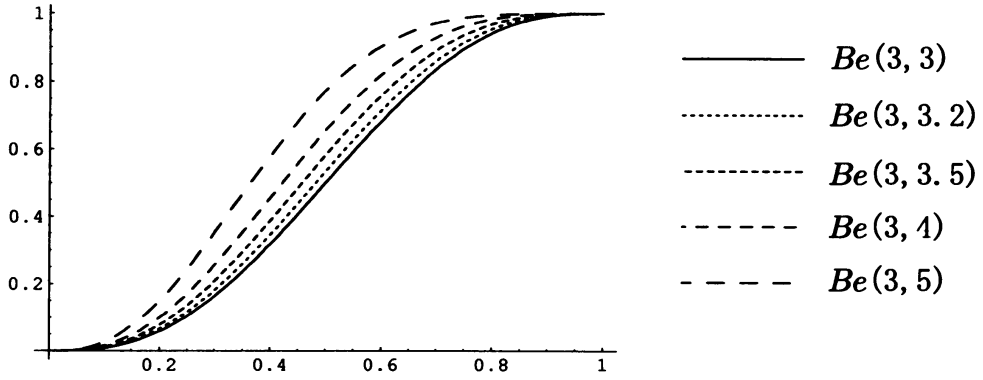

Figure 2 *The c.d.f.'s of the beta distributions*

As is seen from Table 1, the values of approximate power are comparatively stable except for the cases when $(a, b) = (4, 4.5)$ and $m = 100$, which are due to the first sample from the distribution $Be(4, 4)$. Table 2 shows a similar tendency to the result in Table 1. (ii) Let F be the c.d.f. of the standard normal distribution $N(0, 1)$ and G the c.d.f. of $N(\mu, \sigma^2)$. For $\alpha = 0.05$, $(\mu, \sigma^2) = (0, 2), (0, 3), (0, 4), (0.5, 1)$ and $N = 500$, we have the values of the approximate power k_N/N is three simulation results as in Table 3. When $(\mu, \sigma^2) = (0, 2)$, they seems to be unstable.

Table 3 *Comparison of the approximate power k_N/N of the test in three simulation results in case of $N(0, 1)$*

(μ, σ^2)	m	k_N/N		
$N(0, 2)$	50	0.940	0.280	0.536
$N(0, 3)$	50	0.944	0.876	0.780
$N(0, 4)$	50	0.768	0.776	0.984
$N(0.5, 1)$	50	0.988	0.840	1.000
$t(3)$	50	0.322	0.378	0.168
$t(4)$	50	0.180	0.102	0.438

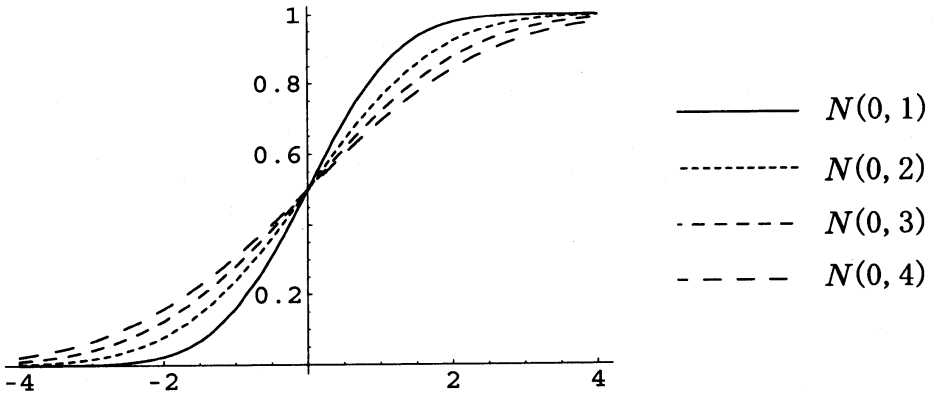

Figure 3 *The c.d.f.'s of the Normal distributions*

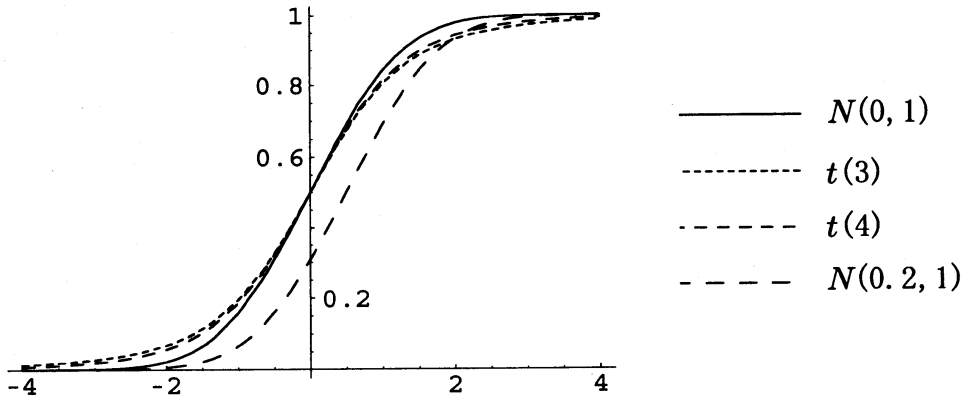

Figure 4 *The c.d.f.'s of the Normal distributions and t-distributions*

3. Remarks

The two-sample problem in this paper may be applied to the following[1]. Suppose that a drug is admitted as a marketable one in some countries after testing its efficacy by many data. When those date are available, the problem is how to test the efficacy of the drug by only *small* data in another country. In the problem the size of small data is important, and, according to our simulation study, the result seems to be comparatively stable for the size 50, though it may depend on the population distribution and the first sample from it.

References

Akahira, M. and Takeuchi, K. (1991). Bootstrap method and empirical process. *Ann. Inst. Statist. Math.*, **43**, 297–310.

Birnbaum, Z. W. (1952). Numerical tabulation of the distribution of Kolmogorov's statistic for finite sample size. *J. Amer. Statist. Assoc.*, **47**, 425–441.

Miller, L. H. (1956). Table of percentage points of Kolmogorov statistics. *J. Amer. Statist. Assoc.*, **51**, 111–121.

[1]Such a medical problem was brought by Prof. M. Takeuchi of Kitasato University to the first author.