International Journal of Pure and Applied Mathematics
Volume 13 No. 2 2004, 235-243

ON SMALL SAMPLE PROPERTIES OF PERMUTATION TESTS: INDEPENDENCE BETWEEN TWO SAMPLES

Hisashi Tanizaki

Graduate School of Economics
Kobe University
2-1, Rokkodai-cho, Nada-ku
Kobe 657-8501, JAPAN
e-mail: tanizaki@kobe-u.ac.jp

Abstract: In this paper, we consider a nonparametric permutation test on the correlation coefficient. Because the permutation test is very computer-intensive, there are few studies on small-sample properties, although we have numerous studies on asymptotic properties with regard to various aspects. In this paper, we aim to compare the permutation test with the t test through Monte Carlo experiments, where an independence test between two samples is taken. We obtain the results through Monte Carlo experiments that the nonparametric test performs better than the t test when the underlying sample is not Gaussian and that the nonparametric test is as good as the t test even under the Gaussian population.

AMS Subject Classification: 62G10
Key Words: nonparametric permutation test, correlation coefficient, independence between two samples

1. Introduction

Suppose that we have two samples. When both samples are independently normally distributed, we can utilize the t test for testing the independence between the two samples. Since the distribution of the two samples is not known

Received: March 23, 2004

in general, we need to check whether the normality assumption is plausible before testing the hypothesis. In the case where the normality assumption is rejected, we can no longer test the independence between the two samples using the t test. In order to improve this problem, in this paper we introduce the permutation test suggested by Fisher [1], which can be applicable to any distribution.

The nonparametric tests based on Spearman rank correlation coefficient and Kendall rank correlation coefficient are very famous. See, for example, Conover [2], Gibbons and Chakraborti [3], Hogg and Craig [4], Hollander and Wolfe [5], Randles and Wolfe [6] and Sprent [7] for the rank correlation tests. In this paper, the permutation test proposed by Fisher [1] is utilized, where we compute the correlation coefficient for each of all the possible combinations and all the possible correlation coefficients are compared with the correlation coefficient based on the original data. Because the permutation test is very computer-intensive, there are few studies on small-sample properties, although we have numerous studies on asymptotic properties with regard to various aspects. In this paper, we aim to compare the permutation test with the t test through Monte Carlo experiments, where an independence test between two samples is taken.

The outline of this paper is as follows. In Section 2, we introduce the nonparametric test based on the permutation test, where we consider testing whether X is correlated with Y for the sample size n. In Section 3, we compare the sample powers of the nonparametric tests and the conventional t test when the underlying data are non-Gaussian.

2. Nonparametric Permutation Test on Correlation Coefficient

Let (X_1, Y_1), (X_2, Y_2), $\cdots$, (X_n, Y_n) be a random sample, where the sample size is n. Consider testing if there is a correlation between X and Y, i.e., if the correlation coefficient ρ is zero or not. The correlation coefficient ρ is defined as: $\rho = \mathrm{Cov}(X, Y)/\sqrt{\mathrm{V}(X)\mathrm{V}(Y)}$, where $\mathrm{Cov}(X, Y)$, $\mathrm{V}(X)$ and $\mathrm{V}(Y)$ represent the covariance between X and Y, the variance of X and the variance of Y, respectively. Then, the sample correlation coefficient $\widehat{\rho}$ is written as: $\widehat{\rho} = S_{XY}/(S_X S_Y)$, where S_{XY}, S_X and S_Y denote the sample covariance between X and Y, the sample variance of X and the sample variance of Y, which are given by: $S_{XY} = (1/n)\sum_{i=1}^n (X_i - \overline{X})(Y_i - \overline{Y})$, $S_X^2 = (1/n)\sum_{i=1}^n (X_i - \overline{X})^2$ and $S_Y^2 = (1/n)\sum_{i=1}^n (Y_i - \overline{Y})^2$. $\overline{X}$ and $\overline{Y}$ represent the sample means of X and Y.

If X is independent of Y, we have $\rho = 0$ and the joint density of X and Y

is represented as a product of the two marginal densities of X and Y, i.e.,

$$f_{xy}(x, y) = f_x(x) f_y(y), \tag{1}$$

where $f_{xy}(x, y)$, $f_x(x)$ and $f_y(y)$ denote the joint density of X and Y, the marginal density of X and the marginal density of Y. Equation (1) implies that for all i and j we consider randomly taking n pairs of X_i and Y_j. Accordingly, for fixed X_1, the possible combinations are given by (X_1, Y_j), $j = 1, 2, \cdots, n$, where we have n combinations. Similarly, for fixed X_2, the possible combinations are (X_2, Y_j), $j = 2, 3, \cdots, n$, i.e., $n - 1$ combinations. Moreover, we have $n - 2$ combinations for X_3, $n - 3$ combinations for X_4 and so on. Therefore, total number of all the possible combinations between X and Y are given by $n!$. For each combination, we can compute the correlation coefficient. Thus, $n!$ correlation coefficients are obtained. The n correlation coefficients are compared with the correlation coefficient obtained from the original pairs of data. If the correlation coefficient obtained from the original data is in the tail of the empirical distribution constructed from the n correlation coefficients, the hypothesis that X is correlated with Y is rejected. The testing procedure above is distribution-free or nonparametric, which can be applicable to almost all the cases. The nonparametric test discussed above is known as a permutation test, which is developed by Fisher [1]. Also, see Stuart and Ord [8].

The order of X_i, $i = 1, 2, \cdots, n$, is fixed and we permute Y_j, $j = 1, 2, \cdots, n$, randomly. Based on the $n!$ correlation coefficients, we can test if X is correlated with Y. Let the $n!$ correlation coefficients be $\widehat{\rho}^{(i)}$, $i = 1, 2, \cdots, n!$. Suppose that $\widehat{\rho}^{(1)}$ is the correlation coefficient obtained from the original data. We can obtain the following three numbers of the combinations:

(i) the number of the combinations less than $\widehat{\rho}^{(1)}$ out of $\widehat{\rho}^{(1)}$, $\widehat{\rho}^{(2)}$, $\cdots$, $\widehat{\rho}^{(n!)}$,

(ii) the number of the combinations equal to $\widehat{\rho}^{(1)}$ out of $\widehat{\rho}^{(1)}$, $\widehat{\rho}^{(2)}$, $\cdots$, $\widehat{\rho}^{(n!)}$,

(iii) the number of the combinations greater than $\widehat{\rho}^{(1)}$ out of $\widehat{\rho}^{(1)}$, $\widehat{\rho}^{(2)}$, $\cdots$, $\widehat{\rho}^{(n!)}$.

Therefore, the estimator of the correlation coefficient ρ, denoted by $\widehat{\rho}$, is distributed as:

$$P(\widehat{\rho} < \widehat{\rho}^{(1)}) = \frac{\text{the number of the combinations less than } \widehat{\rho}^{(1)}}{\text{the number of all the possible combinations (i.e., } n!)},$$

$$P(\widehat{\rho} = \widehat{\rho}^{(1)}) = \frac{\text{the number of the combinations equal to } \widehat{\rho}^{(1)}}{\text{the number of all the possible combinations (i.e., } n!)},$$

$$P(\widehat{\rho} > \widehat{\rho}^{(1)}) = \frac{\text{the number of the combinations greater than } \widehat{\rho}^{(1)}}{\text{the number of all the possible combinations (i.e., } n!)},$$

which are correspond to (i) – (iii), respectively. Thus, the above three probabilities can be computed. The null hypothesis $H_0 : \rho = 0$ is rejected by the two-sided test if $P(\widehat{\rho} < \widehat{\rho}^{(1)})$ or $P(\widehat{\rho} > \widehat{\rho}^{(1)})$ is small enough.

Note as follows. We have $S_{XY} = (1/n)\sum_{i=1}^{n} X_i Y_i - \overline{XY}$. $\overline{X}$, $\overline{Y}$, S_X and S_Y take the same values without depending on the order of X and Y.

Therefore, $\widehat{\rho}$ depends only on $\sum_{i=1}^{n} X_i Y_i$. $\widehat{\rho}$ is a monotone function of $\sum_{i=1}^{n} X_i Y_i$, which implies that we have one-to-one correspondence between $\widehat{\rho}$ and $\sum_{i=1}^{n} X_i Y_i$. Therefore, for $\sum_{i=1}^{n} X_i Y_i$, we may compute the $n!$ combinations by changing the order of Y_i, $i = 1, 2, \cdots, n$. Thus, by utilizing $\sum_{i=1}^{n} X_i Y_i$ rather than $\widehat{\rho}$, computational burden can be reduced.

As for a special case, suppose that (X_i, Y_i), $i = 1, 2, \cdots, n$, are normally distributed, i.e.,

$$\begin{pmatrix} X_i \\ Y_i \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix} \right).$$

And the null hypothesis $H_0 : \rho = 0$, the sample correlation coefficient $\widehat{\rho}$ is distributed as the following t distribution:

$$\widehat{\rho}\sqrt{(n-2)/(1-\widehat{\rho}^2)} \sim t(n-2).$$

Note that we cannot use the t distribution in the case of testing the null hypothesis $H_0 : \rho = \rho_0$. For example, see Lehmann [9], Stuart and Ord [8, 10] and Hogg and Craig [4]. Generally, it is plausible to consider that (X, Y) is non-Gaussian and that the distribution of (X, Y) is not known. If the underlying distribution is not Gaussian but the t distribution is applied to the null hypothesis $H_0 : \rho = 0$, the realistic testing results cannot be obtained. However, the nonparametric permutation test can be applied even in the non-Gaussian cases, because it is distribution-free.

n	ρ	α	Nonparametric Permutation Test					Parametric t Test				
			N	X	U	L	C	N	X	U	L	C
6	.0	.10	.1008	.0986	.1040	.1035	.1030	.0981	.1019	.1019	.0987	.1180
		.05	.0499	.0509	.0512	.0518	.0515	.0474	.0646	.0507	.0483	.0748
		.01	.0112	.0096	.0113	.0101	.0098	.0099	.0225	.0115	.0093	.0269
	.3	.10	.1491	.1830	.1479	.1621	.2605	.1496	.2284	.1488	.1585	.3055
		.05	.0821	.1108	.0798	.0926	.1639	.0795	.1638	.0794	.0914	.2375
		.01	.0189	.0323	.0191	.0205	.0475	.0203	.0692	.0188	.0225	.1389
	.6	.10	.2876	.3199	.2605	.3014	.3834	.2932	.3814	.2646	.3081	.4384
		.05	.1776	.2145	.1528	.1836	.2696	.1787	.2917	.1568	.1939	.3633
		.01	.0460	.0755	.0386	.0501	.0983	.0481	.1579	.0389	.0564	.2405
	.9	.10	.4674	.4469	.4365	.4682	.4802	.4792	.5144	.4475	.4810	.5325
		.05	.3117	.3206	.2786	.3240	.3636	.3300	.4208	.2887	.3410	.4549
		.01	.0894	.1269	.0745	.1009	.1572	.1060	.2501	.0843	.1240	.3096
8	.0	.10	.1018	.0994	.1040	.1014	.1007	.1020	.0986	.1057	.1025	.1183
		.05	.0523	.0489	.0523	.0549	.0494	.0515	.0609	.0545	.0549	.0791
		.01	.0103	.0111	.0110	.0100	.0093	.0100	.0225	.0126	.0113	.0345
	.3	.10	.1869	.2025	.1748	.1971	.3120	.1866	.2514	.1779	.1978	.3471
		.05	.1066	.1236	.0940	.1096	.2107	.1083	.1821	.0998	.1147	.2780
		.01	.0257	.0385	.0211	.0251	.0749	.0248	.0891	.0250	.0254	.1814
	.6	.10	.3932	.3997	.3605	.4002	.4527	.3996	.4592	.3657	.4049	.4963
		.05	.2602	.2833	.2212	.2611	.3455	.2693	.3695	.2359	.2742	.4205
		.01	.0815	.1148	.0627	.0850	.1615	.0865	.2149	.0714	.0959	.2920
	.9	.10	.6197	.5445	.6256	.6209	.5647	.6289	.6023	.6309	.6313	.5979
		.05	.4653	.4163	.4511	.4850	.4684	.4806	.5125	.4664	.4980	.5315
		.01	.1966	.2034	.1562	.2119	.2646	.2199	.3422	.1795	.2395	.4096
10	.0	.10	.1046	.0987	.0999	.0984	.0992	.1033	.0943	.1006	.0950	.1058
		.05	.0524	.0486	.0520	.0489	.0477	.0510	.0570	.0526	.0479	.0733
		.01	.0106	.0083	.0104	.0102	.0090	.0097	.0220	.0114	.0098	.0348
	.3	.10	.2162	.2243	.2046	.2228	.3468	.2151	.2765	.2065	.2231	.3764
		.05	.1277	.1429	.1179	.1348	.2513	.1284	.2046	.1218	.1348	.3133
		.01	.0338	.0484	.0285	.0351	.1022	.0355	.1037	.0306	.0367	.2149
	.6	.10	.4766	.4501	.4640	.4920	.5089	.4775	.5005	.4692	.4954	.5378
		.05	.3387	.3335	.3118	.3579	.4081	.3416	.4149	.3198	.3599	.4670
		.01	.1325	.1522	.1020	.1424	.2281	.1375	.2604	.1106	.1504	.3610
	.9	.10	.7336	.6128	.7539	.7276	.6075	.7358	.6663	.7558	.7348	.6343
		.05	.6008	.4958	.6038	.6032	.5258	.6116	.5806	.6106	.6129	.5734
		.01	.3116	.2812	.2719	.3286	.3318	.3295	.4128	.2925	.3482	.4553

Table 1: Empirical sizes and sample powers ($H_0 : \rho = 0$ and $H_1 : \rho \neq 0$)

n	8	9	10	11	12
Seconds	0.0066	0.0727	0.6714	8.3445	91.816

Table 2: Computational time for N and $\rho = 0$

3. Monte Carlo Experiments

Each value in Table 1 represents the rejection rates of the null hypothesis $H_0 : \rho = 0$ against the alternative hypothesis $H_1 : \rho \neq 0$ (i.e., the two-sided test is chosen) by the significance level $\alpha = 0.01, 0.05, 0.10$, where the experiment is repeated 10^4 times. That is, in Table 1 the number which the correlation coefficient obtained from the original observed data is less than $\alpha/2$ or greater than $1-\alpha/2$ is divided by 10^4. In other words, the ratios of $P(\widehat{\rho} < \widehat{\rho}^{(1)}) \leq \alpha/2$ or $P(\widehat{\rho} > \widehat{\rho}^{(1)}) \leq \alpha/2$ based on 10^4 simulation runs are shown in Table 1. $\alpha = 0.10, 0.05, 0.01$ is examined. N, X, U, L and C indicate the underlying distributions of $(X,\, Y)$, which denote the standard normal distribution $N(0,1)$, the chi-squared distribution $\chi^2(1)-1$, the uniform distribution $U(-1,1)$, the logistic distribution $e^{-x}/(1+e^{-1})^2$, and the Cauchy distribution $(\pi(1+x^2))^{-1}$, respectively. The standard error of the empirical power, denoted by $\widehat{p}$, is obtained by $\sqrt{\widehat{p}(1-\widehat{p})/10^4}$. For example, when $\widehat{p} = 0.5$ the standard error takes the maximum value, which is 0.005.

In this paper, the random draws of $(X_i,\, Y_i)$ are obtained as follows. Let u_i and v_i be the random variables which are mutually independently distributed. Each of u_i and v_i is generated as the standard normal random variable $N(0,1)$, the chi-squared random variable $\chi^2(1)-1$, the uniform random variable $U(-1,1)$, the logistic distribution $e^{-x}/(1+e^{-1})^2$, or the Cauchy distribution $(\pi(1+x^2))^{-1}$. Denote the correlation coefficient between X and Y by ρ. Given the random draws of $(u_i,\, v_i)$ and ρ, $(X_i,\, Y_i)$ is transformed into:

$$\begin{pmatrix} X_i \\ Y_i \end{pmatrix} = \begin{pmatrix} 1 & \rho \\ 0 & \sqrt{1-\rho^2} \end{pmatrix} \begin{pmatrix} u_i \\ v_i \end{pmatrix}. \tag{2}$$

In the case of the Cauchy distribution the correlation coefficient does not exist because the Cauchy random variable has neither mean nor variance. Even in the case of the Cauchy distribution, however, we can obtain the random draws of $(X_i,\, Y_i)$ given $(u_i,\, v_i)$ and ρ, utilizing the formula (2).

As it is easily expected, for the normal sample (i.e., N), the t test performs better than the nonparametric test, but for the other samples (i.e., X, U, L and C), the nonparametric test is superior to the t test.

Each value in the case of $\rho = 0$ represents the empirical size, which should be theoretically equal to the significance level α. In the case of N and L, each value is quite close to α for the t test. However, for all $n = 6, 8, 10$, the t tests of X and C are over-rejected especially in the case of $\alpha = 0.05, 0.01$. Taking an example of $n = 6$, X is 0.0646 and C is 0.0748 when $\alpha = 0.05$, while X is 0.0225 and C is 0.0269 when $\alpha = 0.01$. The t test of U is also slightly over-rejected in the case of $\alpha = 0.05, 0.01$. Thus, we often have the case where the t test over-rejects the null hypothesis $H_0 : \rho = 0$ depending on the underlying distribution. However, the nonparametric test rejects the null hypothesis with probability α for all the underlying distributions. Accordingly, through a comparison of the empirical size, we can conclude that the nonparametric test is more robust than the t test.

Next, we examine the cases of $\rho = 0.3, 0.6, 0.9$ to compare the sample powers of the two tests (note that each value of $\rho = 0.3, 0.6, 0.9$ in Table 1 corresponds to the sample power). As for X and C, it is not meaningful to compare the sample powers because the empirical sizes are already over-estimated. Regarding the sample powers of N, U and L, the nonparametric test is close to the t test. Especially, for N, it is expected that the t test is better than the nonparametric test, but we can see that the nonparametric test is as good as the t test in the sense of the sample power.

Thus, the permutation-based nonparametric test introduced in this paper is useful because it gives us the correct empirical size and is powerful even though it does not need to assume the distribution function.

In Table 2, computational time is shown for N and $\rho = 0$. Note that *Pentium* III-S 1.40GHz CPU, *Microsoft Windows 2000* operating system and *Open Watcom* F77/32 Compiler Version 1.2 (`http://www.openwatcom.org`) are utilized for all the computations in Tables 1 and 2.. Each value indicates the average second from 100 simulation runs, i.e., computational time per one set of two samples. For example, 8.3445 in the case of $n = 11$ implies that it takes 8.3445 seconds to obtain 11! correlation coefficients. In the case of $n = 12$, it takes 91.816 seconds to test independence between two samples. Thus, as n increases, computational time extraordinarily increases.

4. Summary

When the two samples are independently normally distributed, we can utilize the t test for independence between two samples. Since the underlying distribution is not known in general, we need to check whether the normality assumption holds before testing the hypothesis. In the case where the normality assumption is rejected, we cannot test the independence between two samples using the t test. In order to improve this problem, in this paper we have shown the independence test between two samples, which can be applicable to any distribution.

In Section 3, We have tested whether the correlation coefficient between two samples is zero and examined the sample powers of the permutation test and the t test. For each of the cases where the underlying samples are normal, chi-squared, uniform, logistic and Cauchy, 10^4 simulation runs are performed and the nonparametric permutation test is compared with the parametric t test with respect to the empirical sizes and the sample powers. As it is easily expected, the t test is sometimes a biased test under the non-Gaussian assumption. That is, we have the cases where the empirical sizes are over-estimated. However, the nonparametric test gives us the correct empirical sizes without depending the underlying distribution. Specifically, even when the sample is normal, the nonparametric test is very close to t test (theoretically, the t test should be better than any other test when the sample is normal). Thus, we have obtained the results through Monte Carlo experiments that the nonparametric test performs better than the t test when the underlying sample is not Gaussian and that the nonparametric test is as good as the t test even under the Gaussian population.

However, we have the problem of the computational cost. As n is large, an amount of computation on $P(\widehat{\rho} < \rho^{(1)})$, $P(\widehat{\rho} = \rho^{(1)})$ and $P(\widehat{\rho} > \rho^{(1)})$ extremely increases.

Acknowledgments

This research was partially supported by Japan Society for the Promotion of Science, Grants-in-Aid for Scientific Research (C) #14530033, 2002 – 2005, and Grants-in-Aid for the 21st Century COE Program.

References

[1] W.J. Conover, *Practical Nonparametric Statistics*, Second Edition, John Wiley and Sons (1980).

[2] R.A. Fisher, *The Design of Experiments*, Eighth Edition, New York, Hafner (1966).

[3] J.D. Gibbons, S. Chakraborti, *Nonparametric Statistical Inference*, Third Edition, Revised and Expanded, Marcel Dekker (1992).

[4] R.V. Hogg, A.T. Craig, *Introduction to Mathematical Statistics*, Fifth Edition, Prentice Hall (1995).

[5] M. Hollander, D.A. Wolfe, *Nonparametric Statistical Methods*, John Wiley and Sons (1973).

[6] E.L. Lehmann, *Testing Statistical Hypotheses*, Second Edition, John Wiley and Sons (1986).

[7] R.H. Randles, D.A. Wolfe, *Introduction to the Theory of Nonparametric Statistics*, John Wiley and Sons (1979).

[8] P. Sprent, *Applied Nonparametric Statistical Methods*, Chapman and Hall (1989).

[9] A. Stuart, J.K. Ord, *Kendall's Advanced Theory of Statistics*, Volume **2**, Fifth Edition, Edward Arnold (1991).

[10] A. Stuart, J.K. Ord, *Kendall's Advanced Theory of Statistics*, Volume **1**, Sixth Edition, Edward Arnold (1994).