USING NONPARAMETRIC TEST TO COMPARE THE PERFORMANCE OF METAHEURISTICS[1]

Juan G. Villegas
Assistant professor- Department of Industrial Engineering-Universidad de Antioquia (Medellin-Colombia)

FRIEDMAN TEST (based on Conover 1998)

PROCEDURE

The Friedman test can be used to compare the performance of k ($k \geq 2$) metaheuristics using a test bed with b instances, and (only) a single run for each metaheuristic. The data of the experiment will be presented in a table b x k, where entry X_{bk} reports the objective function found by Metaheuristic k in problem b[2]

	Metaheuristic			
Instance	*Metaheuristic* 1	*Metaheuristic* 2	…	*Metaheuristic k*
Instance 1	X_{11}	X_{12}		X_{1k}
Instance 2	X_{21}	X_{22}		X_{2k}
…				
Instance b	X_{b1}	X_{b2}		X_{bk}

The assumptions of Freidman test are:

- The results over instances are mutually independent (i.e. the results within one instance do not influence the results within other instance)
- Within each instance the observations (objective functions) can be ranked

Hypotheses:

H_0: Each ranking of the metaheuristics within each problem is equally likely, (i.e., there is no difference between them)

H_1: At least one of the metaheuristics tends to yield larger objective functions than at least one of the other metaheuristics

The procedure to perform the Friedman test is as follows:

- Rank the results of the metaheuristics within each instance, giving 1 to the best and k to the worst. Let $R(X_{ij})$ be the rank, from 1, to k, assigned to X_{ij} in problem i. In case of ties use average ranks.
- Calculate the total summation of squared ranks A_2:

[1] This document is based on pages 294-302 of: Conover, W. (1998). Practical nonparametric statistics. New York: Wiley.
[2] Several runs of each metaheuristic for each problem can be used and in that case we use the average of the runs as the entries of the table

$$A_2 = \sum_{i=1}^{b} \sum_{j=1}^{k} [R(X_{ij})]^2$$

In the absence of ties A_2 simplifies to:

$$A_2 = \frac{bk(k+1)(2k+1)}{6}$$

Compute the summation of the rank for each metaheuristic $R_j = \sum_1^b R(X_{ij})$ for $j = 1, \dots k$, and calculate B_2 :

$$B_2 = \frac{1}{b} \sum_{j=1}^{k} R_j^2$$

The test statistic is given by:

$$T_2 = \frac{(b-1)[B_2 - bk(k+1)^2/4]}{A_2 - B_2}$$

Reject the null hypothesis at the level of significance α if T_2 is greater than the $1-\alpha$ quantile of the F distribution with $k_1 = k-1$ and $k_2 = (k-1)(b-1)$ degrees of freedom.

EXAMPLE

Example: Using the data from Villegas et al. (2011) for the average objective function of 4 different metaheuristics on 21 test problems of the truck and trailer routing problem we have the following table:

	Metaheuristic 1			Metaheuristic 2			Metaheuristic 3			Metaheuristic 4		
Instance (b)	X_{b1}	$R(X_{b1})$	$R(X_{b1})^2$	X_{b2}	$R(X_{b2})$	$R(X_{b2})^2$	X_{b3}	$R(X_{b3})$	$R(X_{b3})^2$	X_{b4}	$R(X_{b4})$	$R(X_{b4})^2$
1	565.02	1	1	567.98	3	9	568.86	4	16	565.99	2	4
2	662.84	4	16	619.35	3	9	617.48	2	4	614.23	1	1
3	664.73	4	16	629.59	3	9	620.50	2	4	618.04	1	1
4	857.84	4	16	809.13	2	4	817.71	3	9	803.51	1	1
5	949.98	4	16	858.98	3	9	858.95	2	4	841.63	1	1
6	1084.82	4	16	949.89	2	4	942.60	1	1	961.47	3	9
7	837.80	3	9	832.91	2	4	838.50	4	16	830.48	1	1
8	906.16	4	16	881.26	2	4	882.70	3	9	876.21	1	1
9	1000.27	4	16	955.95	3	9	921.97	2	4	918.45	1	1
10	1076.88	4	16	1052.65	2	4	1074.38	3	9	1050.11	1	1
11	1170.17	4	16	1107.47	2	4	1108.88	3	9	1100.95	1	1
12	1217.01	4	16	1184.58	3	9	1166.59	2	4	1158.88	1	1
13	1364.50	4	16	1296.33	1	1	1340.98	3	9	1305.83	2	4
14	1464.20	4	16	1384.13	3	9	1367.91	2	4	1354.04	1	1
15	1544.21	4	16	1488.71	3	9	1454.91	2	4	1437.52	1	1
16	1064.89	4	16	1003.00	1	1	1007.26	3	9	1003.07	2	4
17	1104.67	4	16	1042.79	3	9	1035.23	1	1	1042.61	2	4
18	1202.00	4	16	1141.94	3	9	1110.13	1	1	1118.63	2	4
19	887.22	4	16	813.98	1	1	823.01	3	9	819.81	2	4
20	963.06	4	16	852.89	1	1	859.06	2	4	860.12	3	9
21	952.29	4	16	914.04	2	4	915.38	3	9	909.06	1	1
Average Rank		3.81			2.29			2.43			1.48	
Summation		80	314		48	122		51	139		31	55

Using the values of the table it is possible to calculate:

$$A_2 = 314 + 122 + 139 + 55 = 630$$

$$B_2 = \frac{1}{21}[80^2 + 48^2 + 51^2 + 31^2] = 584.1$$

With the values of A_2 and B_2, calculate the test statistic T_2:

$$T_2 = \frac{(21-1)[584.1 - 21 \times 4 \times (5)^2/4]}{630 - 584.1} = 25.75$$

Using a table for the F distribution with a significance level $\alpha = 0.01$ we found that

$$F_{1-\alpha,k-1,(b-1)(k-1)} = F_{0.99,3,60} = 4.13$$

Since $T_2 > F_{0.99,3,60}$ we reject the null hypothesis, then there exist at least one metaheuristic whose performance is different from at least one of the other metaheuristics.

However, it is necessary to perform paired comparisons to know which metaheuristics are really different

(POST HOC) PAIRED COMPARISONS

The following method is used to know if metaheuristics i and j are considered different after the rejection of the null hypothesis with the Friedman test. Calculate the absolute difference of the summation of the ranks of metaheuristics i and j and declare i and j different if :

$$|R_i - R_j| > t_{1-\frac{\alpha}{2}} \left[\frac{2b(A_2 - B_2)}{(b-1)(k-1)}\right]^{\frac{1}{2}}$$

where $t_{1-\frac{\alpha}{2}}$ is the $1-\frac{\alpha}{2}$ quantile of the t distribution with $(b-1)(k-1)$ degrees of Freedom.

In the Example $t_{1-\frac{\alpha}{2}}$ for $\alpha = 0.01$ and 60 degrees of Freedom is 2.660 and the critical value for the difference is: $2.660\left[\frac{2\times 21\times(630-584.1)}{(20)(3)}\right]^{\frac{1}{2}} = 15.08$

The table summarizes the paired comparisons; underlined values indicated that the metaheuristics are different.

$\lvert R_i - R_j\rvert$	*Metaheuristic* 2	*Metaheuristic* 3	*Metaheuristic* 4
Metaheuristic 1	<u>32</u>	<u>29</u>	<u>49</u>
Metaheuristic 2	-	3	<u>17</u>
Metaheuristic 3	-	-	<u>20</u>

From the above table, we can see that: metaheuristic 1 is outperformed by all the other metaheuristics, metaheuristic 2 and 3 have the same performance; and metaheuristic 4 is the best, outperforming each one of the other metaheuristics.

NOTES AND REFERENCES

- There exists another test statistic for the Friedman test that uses a chi-square distribution. However, Conover (1998) recommends the use of the F approximation because some studies show that it is better.
- If there are only two metaheuristics it is also possible to use a simpler test called the *sign test,* based on the binomial distribution[3].
- See http://www.fon.hum.uva.nl/Service/Statistics.html for a server that performs several nonparametric tests.

[3] See http://www.fon.hum.uva.nl/Service/Statistics/Sign_Test.html for a brief description of the sign test and one online tool for its computations,

For detailed discussions of nonparametric test used for the comparison of metaheuristics see:

- É. D. Taillard, Ph. Waelti, J. Zuber, Few statistical tests for proportions comparisons. *European Journal of Operational Research* 185 (3), 1336-1350, 2008.

- García, S., Molina, D., Lozano, M., & Herrera, F. (2009). A study on the use of non-parametric tests for analyzing the evolutionary algorithms' behavior: a case study on the CEC'2005 Special Session on Real Parameter Optimization. *Journal of Heuristics*, 15 (6), 617-644.

The data used to illustrate the test has been taken from:

- J. G. Villegas, C. Prins, C. Prodhon, A. L. Medaglia, N. Velasco: A GRASP with evolutionary path relinking for the truck and trailer routing problem. *Computers & Operations Research* 38(9): 1319-1334 (2011)