Study and implementation of Streitberg-Röhmel “shift” algorithm in R

Student: Setti BELHOUARI
Department of Chemistry and Biomolecular Sciences
Faculty of Science
University of Ottawa

Professor: Gilles LAMOTHE
Department of Mathematics and Statistics

Introduction

The Wilcoxon signed-rank test is a non-parametric test used when the sample cannot be assumed normal, but symmetric, such that it follows at least an ordinal scale. The test ranks the absolute value of the differences of the dependent samples (allowing tied ranks), and attributes the sign of the difference to the rank. The Wilcoxon statistic, ***W***, is the sum of all signed ranks. Starting with a null hypothesis, H_0: $m_d=0$, where m_d is the median of d and d is the difference of the pairs, one can test the following hypotheses: H_1: $m_d>0$, H_2: $m_d<0$, or H_3: $m_d \neq 0$, respectively, by writing all possible positive and negative permutations of the ranks, determining the sum of each permutation, w, and calculating the following p-values, P(w ≥ ***W***), P(w ≤ ***W***), and 2P(w ≤ |***W***|). Though the Wilcoxon signed-rank test is a robust statistic, listing the 2^n permutations is tedious, where n is the sample size.

Although, the Streitberg-Röhmel “shift” algorithm allows for the exact calculation of the Wilcoxon one-sample test, it has not been implemented in most statistical software. This study implements the algorithm in popular statistical software, R, and compares its speed and accuracy to the available inaccurate variations in R, namely the wilcox.test function, which uses an asymptotic normal approximation and assumes no ties, and the Monte Carlo simulation. A comparative simulation is performed to describe the trade-off between accuracy and speed.

Figure 1. An example of the test statistic as proposed by Wilcoxon.

T_1	T_2	$T_1 - T_2$	$\lvert T_1 - T_2 \rvert$	Rank	Signed rank	*W*
1	5	-4	4	3	-3	-2
2	4	-2	2	1	-1	
4	1	3	3	2	2	

Signed rank permutations	w	Test statistic	P-value
1,2,3	6	P(w≥-2)	0.75
-1,2,3	4		
1,-2,3	2		
1,2,-3	0	P(w≤-2)	0.375
-1,-2,3	0		
-1,2,-3	-2		
1,-2,-3	-4	2P(w≥|-2|)	0.75
-1,-2,-3	-6		

Figure 2. The “shift” algorithm as proposed by Streitberg-Röhmel for the above example.

Ranks	0 : sum of ranks 0	1	2	3	4	5	6
	1						
3	+ 0	0	0	1			
	1	0	0	1			
1	+ 0	1	0	0	1		
	1	1	0	1	1		
2	+ 0	0	1	1	0	1	1
	1	1	1	2	1	1	1
w	0	1	2	3	4	5	6
Occurrence	1	1	1	2	1	1	1
W_+	Sum of positive ranks = 2						
P(w≥2)	6/8 = 0.75						
W_-	Absolute sum of negative ranks = 4						
P(w≥4)	3/8 = 0.375						
P(w≤2)+ P(w≥4)	6/8 = 0.75						

Methodology

Results

Figure 3. Comparative simulation of the Streitberg-Röhmel “shift” algorithm, the Monte Carlo simulation, and the wilcox.test function in R. In terms of speed, the shift algorithm, which calculates the exact p-value, competes effectively against the Monte Carlo simulation and the wilcox.test function when the sample size is less than 100. At samples exceeding 1000, it is outrun by both the Monte Carlo simulation and the wilcox.test. t_{avg} and σ represent average time and the standard deviation of the time in seconds, respectively.

Sample size	t_{avg} shift algorithm (s)	σ shift algorithm (s)	t_{avg} Monte Carlo simulation (s)	σ Monte Carlo simulation (s)	t_{avg} wilcox.test (s)	σ wilcox.test (s)
10	8e-04	0.00367354	0.8901	0.1472612	0.0016	0.00486587
50	0.0023	0.00565953	1.0978	0.0551688	0.0022	0.00612661
100	0.0161	0.00941683	1.3374	0.0638751	0.0012	0.00408989
500	1.666	0.1191468	3.6182	0.2328323	0.0029	0.00640312
1000	13.558	0.3202903	6.6181	0.2686957	0.0046	0.00744271
2000	218.2384	927.7912	12.5483	1.398673	0.0104	0.00827799

Figure 4. Graphical representation of the average time taken by the “shift algorithm”, the Monte Carlo simulation, and the wilcox.test function, depending on the sample size. At sample sizes lower than 500, the “shift” algorithm outdoes the Monte Carlo simulation and the wilcox.test function; however, its performance time increases exponentially and follows $t_{avg} = 0.9778e^{0.002707n} - 1.321$, where *t* is the time in seconds and *n* is the sample size.

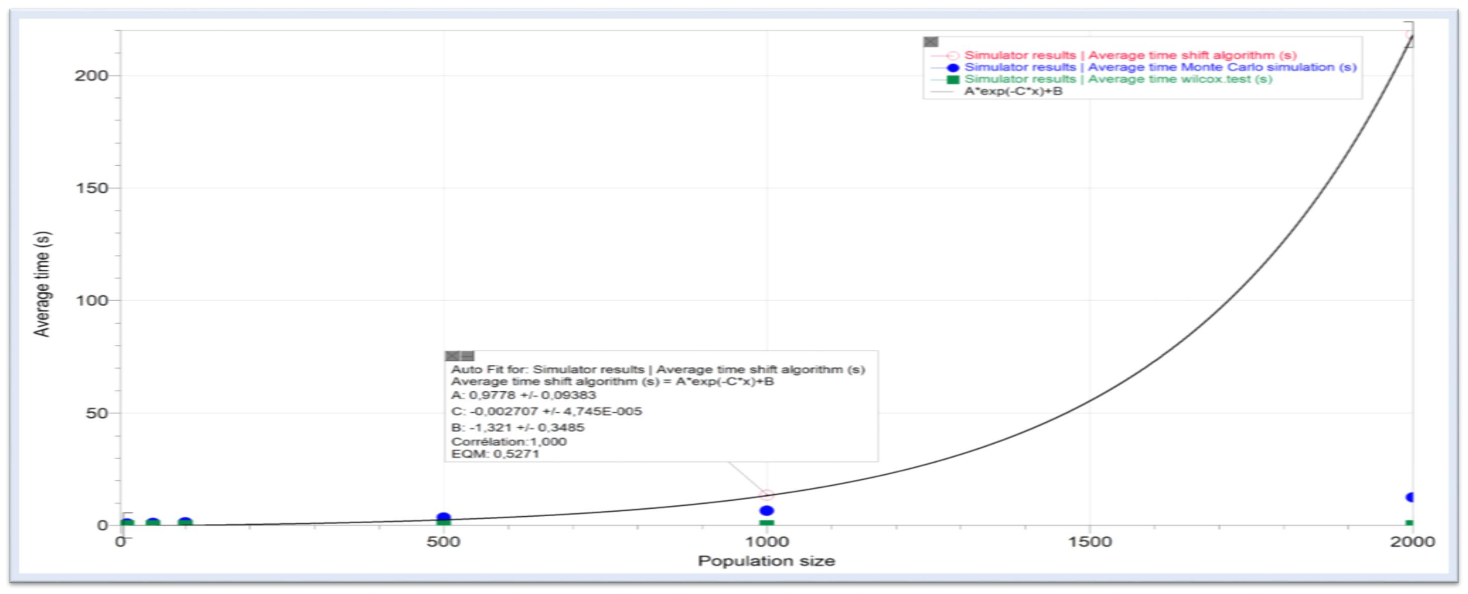


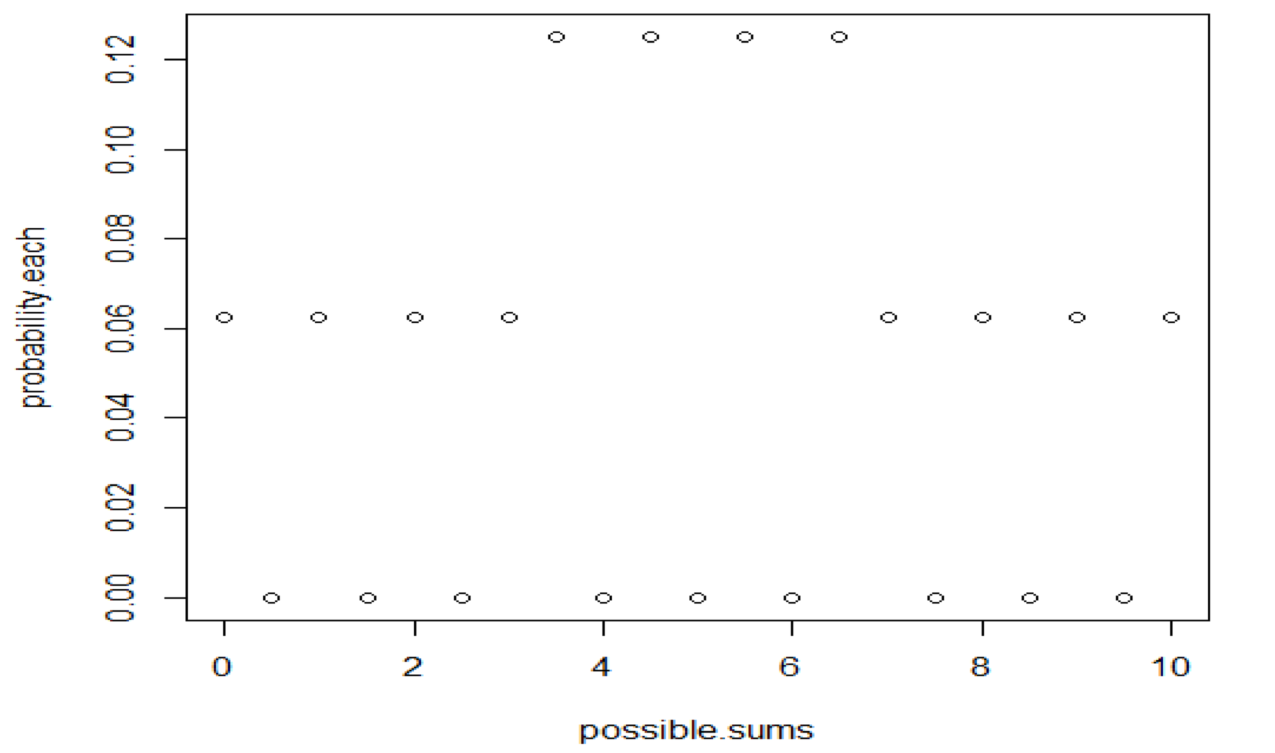
Figure 5. Error in the p-value calculated by the Monte Carlo simulation and the wilcox.test function relative to that calculated by the exact Streitberg-Röhmel “shift” algorithm. The p-values calculated by the Monte Carlo simulation and the wilcox.test function are surprisingly similar, indicating that even in the absence of impractical assumptions (i.e. normality and absence of ties), the Monte Carlo simulation cannot accurately represent the exact Wilcoxon distribution.

Sample size	Relative error: Monte Carlo simulation vs. shift algorithm	Wilcox.test vs. shift algorithm	Monte Carlo simulation vs. shift algorithm
10	0.05808661	0.07679003	0.07562106
50	0.0780207	0.07554451	0.07213257
100	0.09065826	0.104102	0.07798579
500	0.1625768	0.1599788	0.06048961
1000	0.2226726	0.2176888	0.06642389
2000	0.281572	0.2798163	0.06882044

Figure 6. Example of the output of the implemented Streitberg-Röhmel “shift” algorithm in R.

```
> shift.algorithm(x=c(1,2,3,4),y=c(5,4,7,1))
Exact Wilcoxon statistic as per Streitberg-Röhmel shift algorithm

Difference :  -4 -2 -4 3
Wilcoxon statistic or sum of ranks :  10
Sum of positive sign ranks :  2
Sum of negative sign ranks :  8
Sum of zero sign ranks :  0
Shift values :  7 2 7 4
Wilcoxon distribution :
Right-sided pvalue :  0.875
> #Wilcoxon distribution
```



Conclusion

The Streitberg-Röhmel “shift” algorithm, which calculates the exact p-value of the Wilcoxon signed rank test comparing dependent paired samples, can be implemented in R. Given the exactitude of the method, it is strongly recommended to use the algorithm when comparing 500 dependent paired samples or less. For sample sizes larger than 500 pairs, use $t_{avg} = 0.9778e^{0.002707n} - 1.321$, to approximate the processing time for the computation of the p-value. Otherwise, perform a Monte Carlo simulation with more than 5000 permutations to obtain a sufficiently accurate p-value. Both the Streitberg-Röhmel “shift” algorithm and the Monte Carlo simulator are available in R.

Bibliography

[1] Streitberg, B., & Röhmel, J. (1987). Exakte Verteuilungen für Rang-und Randomisierungstests im allgemeinen c-Stickprobenproblem. EDV in Medizin und Biologie. *EDV in Medizin und Biologie,*18(1), 1st ser., 12-19.

[2] Hogg, R. V., & Tanis, E. A. (2006). *Probability and statistical inference*(7th ed.). Upper Saddle River, NJ: Prentice Hall.

[3] Lowry, R. (2017). *Concepts and Applications of Inferential Statistics*. Poughkeepsie, NY: Vassar College.

Acknowledgements

Many thanks to Dr. Gilles Lamothe for teaching me repeatedly the similarities and the differences between the Wilcoxon statistic, the Streitberg-Röhmel algorithm, and the Walsh averages, as well as debugging the R codes.
Many thanks to the University of Ottawa for making this project possible.