

Permutation Tests for Survival Data Analysis

Rosa Arboretti¹, Dario Basso², Federico Campigotto², Luigi Salmaso²

¹Dipartimento Territorio e Sistemi Agro-Forestali, Università di Padova

²Dipartimento di Tecnica e Gestione dei sistemi industriali, Università di Padova

E-mail: rosa.arboretti@unipd.it, basso@gest.unipd.it, federico.campigotto@gmail.com,
salmaso@gest.unipd.it

Abstract: This work moves from the need to find out robust methods in the framework of survival analysis. The attention has been paid in particular to the problem of hypotheses testing in two independent samples for both one and two-sided alternatives on right-censored survival data and to the search for solutions within the permutation framework. Through a comparative simulation study and taking into account for a wide variety of commonly encountered problems in real applications, we have evaluated the behavior of some permutation solutions, including a multidimensional permutation-based combination test in comparison with asymptotic solutions from the recent literature and with traditional asymptotic nonparametric tests.

Keywords: survival analysis; right-censoring data; unequal censoring; nonparametric tests, permutation tests.

1. Introduction

Let us refer to the two-sample hypotheses testing problem within survival analysis in which a typical right-censored survival data set consists of n independent realizations of the random vector (X_{mj}, Δ_{mj}) for $j=1,2$ and $m=1, \dots, n_j$. The variable X_{mj} denotes the time for which subject m in group j is under observation and the variable Δ_{mj} takes the value 1 if the observation on subject m is uncensored and takes the value 0 if the observation is censored. In survival analysis, it is often of interest to test whether or not two survival time distributions are equal. We assume that independent observations are available on the failure times of n individuals and that an individual can fail at most once. The testing problem is usually formulated as testing the null hypothesis $H_0 : S_1(t) = S_2(t) \quad \forall \quad t \in \mathbb{R}^+$ against the alternative hypothesis $H_1 : S_1(t) \leq \neq \geq S_2(t)$ for some $t \in \mathbb{R}^+$.

2. 2. A comparative simulation study

A comparative simulation study has been performed in order to evaluate the behavior of the permutation approaches with respect to some other asymptotic nonparametric methods proposed in the recent literature and looking for possible flexible and robust procedures in terms of power among the investigated methods. In the random censorship model, the asymptotic weighted log-rank (WLR-Mantel [1966]) and weighted Kaplan-Meier (WKM; Pepe and Fleming [1989]) tests are the most commonly used procedures for comparing samples of right-censoring survival data. The former test can be viewed as censored data generalization of such familiar linear rank statistics as

the Wilcoxon test and the Savage (exponential scores) test. The WKM test is based directly on the difference of the Kaplan-Meier estimates of survival functions in the two samples.

In the framework of permutation methods, one can view censored data as a missing data problem. In this sense, one can carry out a multidimensional permutation test based on the theory of permutation testing with missing data (Pesarin, 2001). In fact, if we assume that, once we have fixed an observed time, the data already censored may be considered as missing data, then it is possible to extend the theory of permutation analysis of missing values to the right-censored survival analysis. Moreover, we may opt for the permutation versions of the log-rank test or the Renyi-type tests. Heimann and Neuhaus (1998) prove that, when the number of events is very small, the permutation log-rank test is valid only for equal censoring.

Callegaro et al. (2003) show that the permutation Renyi-type test can be recommended even under the condition of unequal censoring, especially when the (small) sample sizes of the two groups are strongly unbalanced. Hence, we consider here a multidimensional permutation test in case of equal (**EcMPT**; Callegaro et al., 2003) and unequal censoring (**UcMPT**; Bonnini et al., 2005); the permutation conditional (exact) Renyi-type test (**RUPT**; Callegaro et al., 2003); the permutation (**PLR**; Heimann and Neuhaus, 1998) and the asymptotic versions (**LR**) of the log-rank test; the weighted Kaplan-Meier test (**WKM**) and an integrated WKM version (**iWKM**; Lee et al., 2008) and, finally, a modified Fleming-Harrington's test (**mFH**; Gaugler et al., 2007). Lee et al., (2008) show that the integrated WKM test enhances the results of the WKM test, especially when two hazard functions cross. Furthermore, the modified formulation of Fleming-Harrington's test uses the Peto-Peto and Kelbflish-Prentice estimate of the survival function based on the combined sample evaluated at the current event time.

In order to compare the performance of the proposed solutions, Monte Carlo (MC) simulations have been performed to estimate power under different experimental settings (MC=1000 both in H_0 and in H_1 with a number of permutations $B=1000$). We considered different settings which are of particular interest because they present a wide range of characteristics. Indeed, these configurations will present most of the challenges faced in this type of problems and we will investigate if there is a particular test that deals with all of these issues in a 'robust' way. More precisely, considered configurations regard i) the type of hazard model under alternative hypothesis (models with proportional hazard rates, early, middle, late and crossing hazard differences), ii) the type of censoring model (equal or unequal censoring), iii) the sample sizes (balanced, left or right unbalanced sample sizes) and finally, iv) the type of alternative (one sided or two sided alternatives).

Tables 1-3 summarize the power behavior of the methods in some of the investigated configurations. The whole simulation study is available at: www.gest.unipd.it/~salmaso.

Table 1: Models with middle hazard differences under equal censoring: one-sided alternative with right-unbalanced sample sizes ($n_1=10; n_2=50$)

α	Permutation tests					Asymptotic tests		
	EcMPT	UcMPT	PLR	RUPT	LR	WKM	iWKM	mFH
0.01	0.110	0.478	0.024	0.042	0.041	0.061	0.060	0.116
0.025	0.199	0.494	0.041	0.086	0.066	0.113	0.105	0.165
0.05	0.301	0.521	0.072	0.146	0.100	0.165	0.157	0.225
0.10	0.432	0.553	0.137	0.275	0.168	0.245	0.240	0.319
0.20	0.585	0.592	0.245	0.471	0.277	0.387	0.421	0.453
0.30	0.698	0.621	0.353	0.651	0.374	0.490	0.580	0.536
0.40	0.782	0.637	0.449	0.763	0.472	0.581	0.771	0.620
0.50	0.840	0.659	0.540	0.866	0.554	0.678	0.875	0.708
0.60	0.883	0.684	0.650	0.897	0.666	0.750	0.941	0.788
0.70	0.923	0.704	0.748	0.957	0.769	0.825	0.965	0.847
0.80	0.957	0.723	0.854	0.980	0.865	0.887	0.973	0.903
0.90	0.981	0.748	0.925	0.987	0.940	0.944	1.000	0.961
1.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 2: Models with late hazard differences under equal censoring: two-sided alternative with left-unbalanced sample sizes ($n_1=50; n_2=10$).

α	Permutation tests					Asymptotic tests		
	EcMPT	UcMPT	PLR	RUPT	LR	WKM	iWKM	mFH
0.01	0.331	0.867	0.027	0.128	0.070	0.266	0.497	0.088
0.025	0.496	0.878	0.118	0.334	0.164	0.399	0.613	0.189
0.05	0.626	0.890	0.226	0.562	0.289	0.522	0.706	0.298
0.10	0.741	0.903	0.423	0.731	0.464	0.670	0.771	0.464
0.20	0.837	0.916	0.655	0.859	0.675	0.809	0.867	0.653
0.30	0.903	0.929	0.768	0.925	0.789	0.879	0.921	0.776
0.40	0.936	0.935	0.862	0.958	0.867	0.922	0.944	0.844
0.50	0.961	0.939	0.914	0.972	0.920	0.954	0.963	0.905
0.60	0.978	0.949	0.949	0.983	0.951	0.967	0.978	0.934
0.70	0.984	0.956	0.974	0.992	0.975	0.982	0.990	0.959
0.80	0.989	0.959	0.988	0.996	0.989	0.990	0.995	0.980
0.90	0.997	0.966	0.996	0.999	0.996	0.993	0.999	0.995
1.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 3: Models with crossing hazard differences under equal censoring: two-sided alternative with balanced sample sizes ($n_1=10; n_2=10$)

α	Permutation tests					Asymptotic tests		
	EcMPT	UcMPT	PLR	RUPT	LR	WKM	iWKM	mFH
0.01	0.158	0.768	0.071	0.118	0.093	0.206	0.530	0.061
0.025	0.311	0.811	0.130	0.247	0.161	0.380	0.632	0.108
0.05	0.458	0.840	0.210	0.395	0.244	0.535	0.709	0.157
0.10	0.601	0.863	0.340	0.580	0.369	0.677	0.797	0.216
0.20	0.755	0.884	0.500	0.752	0.518	0.816	0.874	0.346
0.30	0.846	0.898	0.605	0.843	0.623	0.868	0.915	0.445
0.40	0.892	0.911	0.699	0.895	0.706	0.902	0.941	0.540
0.50	0.921	0.929	0.773	0.920	0.777	0.928	0.954	0.622
0.60	0.941	0.934	0.819	0.947	0.820	0.953	0.974	0.704
0.70	0.962	0.947	0.874	0.969	0.876	0.973	0.990	0.775
0.80	0.981	0.960	0.928	0.982	0.933	0.986	0.996	0.845
0.90	0.992	0.969	0.966	0.992	0.967	0.996	0.999	0.926
1.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

3. Conclusions

In general, the obtained results gives a strong indication in favour of using a multidimensional permutation test in case of equal censoring. There are a few tests that in almost all the configurations present similar behavior. This is the case of the permutation and the asymptotic versions of the log-rank test, and the case of the weighted Kaplan-Meier (WKM) test and an integrated WKM version. Finally, the modified Fleming-Harrington's test offers another interesting behavior because there are situations in which it appears to be preferable (models with early differences under equal and unequal censoring). However, simulation results indicate that a slightly 'robust' test can be considered the combination-based permutation test, even if, of course, in some situations this is not the best test to be used.

Acknowledgement: The research has been supported by the University of Padova CPDA088513 (Coordinator: Prof. Luigi Salmaso).

References

- Bonnini, S., Salmaso, L., Solari, A., (2005) Multivariate Permutation Tests for Evaluating Effectiveness of Universities through the Analysis of Student Dropouts. *Statistica & Applicazioni* 3, pp. 37-44
- Callegaro, A., Pesarin, F., Salmaso, L., (2003), Test di Permutazione per il Confronto di Curve di Sopravvivenza. *Statistica Applicata* Vol 15, n.2, 241-261
- Gaugler, T., Kim, D., Liao, S., (2007), Comparing Two Survival Time Distributions: An Investigation of Several Weight Functions for the Weighted Logrank Statistic. *Communications in Statistics-Simulation and Computation*, 36: 423-435
- Heimann, G., Neahtaus, G., (1998), Permutational Distribution of the Log-Rank Statistic under Random Censorship with Applications to Carcinogenicity Assays. *Biometrics* 54, 168-184
- Lee, S., H., Lee, E., J., Omolo, B., O., (2008) Using integrated weighted Survival Difference for the Two-Sample Censored Data Problem. *Computational Statistics and Data Analysis* 52, 4410-4416
- Mantel, N., (1966). Evaluation of Survival Data and Two New Rank Order Statistics Arising in its Consideration. *Cancer Chemotherapy Reports* 50, 163—170
- Pepe, M.S., Fleming, T.R., (1991), Weighted Kaplan--Meier Statistics: Large Sample and Optimality Considerations. *Journal of the Royal Statistical Society, Series B* 53, 341--352.
- Pesarin, F. (2001), *Multivariate Permutation Tests with Applications in Biostatistics*, Wiley, Chichester.