Nonparameteric Bayesian analysis of the 2 sample problem with censoring

Kan Shang[1], Cavan Reilly[2]

Second Bayesian Young Statisticians Meeting (BAYSM 2014)
Vienna, September 18–19, 2014

[1] University of Minnesota, Department of Biomedical Informatics and Computational Biology, Minneapolis, U.S.A
shang050@umn.edu

[2] University of Minnesota, Division of Biostatistics and Department of Biomedical Informatics and Computational Biology, Minneapolis, U.S.A
cavanr@biostat.umn.edu

Abstract

Testing for differences between 2 groups is a fundamental problem in statistics and due to developments in Bayesian nonparametrics there is renewed interest in this problem. Here we describe a new class of tests use the connection between the Dirichlet process prior and the Wilcoxon rank sum test but extends this idea to the Dirichlet process mixture model. Given consistency results for this class of models we develop tests that have appropriate frequentist sampling procedures but have the potential to outperform the usual frequentist tests. Extensions to interval and right censoring are considered and an application to a high dimensional data set obtained from a metabolomics investigation demonstrates the practical utility of the method.

Keywords: Bayesian inference; Dirichlet process prior.

1 Introduction

While the problem of testing for differences between 2 groups has a long history, the development of new approaches to inference has created opportunities to improve on these existing approaches. In this paper we use the connection between the Dirichlet process and Wilcoxon's test to explore methods for testing between 2 groups. We consider 2 ways to exploit this connection: one relying on the Dirichlet process prior and one relying on the Dirichlet process mixture prior. In this way we develop a Bayesian counterpart to the Wilcoxon rank sum statistic and the weighted log rank statistic for right and interval censored data.

2 Description of the test

Ferguson (1973) [2] noted that if X is a random variable with iid realizations $X_1, \ldots, X_n$ and Y is another random variable with realizations $Y_1, \ldots, Y_m$ and one seeks to estimate $P(X \leq Y)$, then if one uses independent Dirichlet process priors for the 2 probability distributions that gave rise to the samples, denoted F_X and F_Y, the Bayes estimate of $P(X \leq Y)$ given the data has the simple form

$$\int \hat{F}_X(t)\, d\hat{F}_Y(t), \tag{1}$$

where $\hat{F}_X(t)$ is the Bayes estimate of F_X and $\hat{F}_Y(t)$ is defined analogously. This expression forms the basis for the tests we propose. Briefly, we suppose a random distribution G on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is uncertain and generated by a Dirichlet process, $G \sim D(\alpha G_0)$. However, rather than just using a Dirichlet process prior we also explore the use of a Dirichlet process mixture prior. Our test is based on computing 95% credible intervals for $P(X \leq Y | X_1, \ldots, X_n, Y_1, \ldots Y_m)$ and if the lower bound of this interval exceeds 0.5 or the upper bound is below 0.5 the test rejects the hypothesis that there is no difference between the groups. We call this class of tests extended Ferguson tests (EF). The presence of any kind of censoring poses no problems in the context of our MCMC algorithm as the censored values are sampled from their full conditionals.

3 Simulation studies

3.1 Fully observed data

For the case where there is no censoring we compare our EF test to the Wilcoxon rank sum test. For scenarios with censoring we compare the proposed method to the log-rank test. We apply the computation algorithm in Escobar and West (1995) [1] and also include the acceleration step discussed in Müller and McEachern (1999)[3] [5].

Similarly, we assume the normal/inverse-gamma model and generate data from normal distribution, Student t distribution and binomial distribution. Here we only list the result for normally distributed data.

	$n = 25$		$n = 100$		$n = 500$	
δ	EF	Wilcoxon	EF	Wilcoxon	EF	Wilcoxon
0.00	0.070	0.050	0.057	0.049	0.052	0.050
0.25	0.081	0.059	0.142	0.129	0.534	0.519
0.50	0.168	0.123	0.437	0.413	0.978	0.977
0.75	0.306	0.252	0.765	0.742	1.000	1.000
1.00	0.462	0.415	0.944	0.930	1.000	1.000

Table 1: Simulation results using the Gaussian base measure and data that is distributed according to the Gaussian distribution

For Gaussian distributed data, the result suggests that the proposed EF test obtains an approximately 4% greater power than the Wilcoxon. The result of the other two cases suggests that while with small sample size the proposed EF test has greater power, the Wilcoxon test is comparable to the EF test when the two groups have more data points and larger differences between their means.

3.2 Right censored data

When there is censoring we use the rejection sampling technique described in Kuo and Mallick (1997)[4]. This makes extending the approach from the previous section straightforward. Results are shown in Table 2.

	$n = 25$		$n = 100$		$n = 500$	
δ	EF	Log-rank	EF	Log-rank	EF	Log-rank
0.00	0.079	0.044	0.086	0.060	0.067	0.052
0.25	0.094	0.066	0.201	0.116	0.475	0.363
0.50	0.161	0.109	0.395	0.276	0.904	0.861
0.75	0.244	0.167	0.637	0.532	0.995	0.993
1.00	0.332	0.266	0.814	0.755	1.000	1.000

Table 2: Simulation results using the Gaussian base measure and data that is distributed according to Gaussian distribution with 50% censoring.

Thus the proposed approach has superior power in the presence of censoring too.

4 Discussion

We have demonstrated that using existing methods for the analysis of the DPM model we can develop tests analogous to the Wilcoxon rank sum test and that these Bayesian alternatives have superior power under a wide range of conditions. Future work will explore the connection between these tests and semiparametric alternatives using the well established connections between frequentist semiparametric inference and the log rank test.

References

[1] Escobar, MD., West, M. (1995) " Bayesian density estimation and inference using mixtures." *Journal of the American Statistical Association*, 90,430, 577–588.

[2] Ferguson, TS. (1973), "A Bayesian analysis of some nonparametric problems." *The Annals of Statistics*, 1, 2, 209–230.

[3] Ishwaran, H., James, LF. (2004), "Computational methods for multiplicative intensity models using weighted gamma processes: proportional hazards, marked point processes, and panel count data." *Journal of the American Statistical Association*, 99, 465, 175–190.

[4] Kuo, L., Mallick, B. (1997), "Bayesian semiparametric inference for the accelerated failure-time model." *The Canadian Journal of Statistics*, 25, 4, 457–472.

[5] Müller, P., MacEachern, SN. (1998), "Estimating mixture of Dirichlet process models." *Journal of Computational and Graphical Statistics*, 7,2, 223–238.