Documentation for CIFCOX.c

Author: Xiaolin Fan

Updated: 8/1/2008

Questions or bug reports can be sent to xfan@mcw.edu

Description

This program is for implementation of the Cox-type regression on cumulative incidence function under the competing risks setting. Methods, described in Section 3.3 of Fan (2008), use the mixture of Polya trees (MPT) process priors and are based on the full likelihood.

Input File Format

The program requires some of the GSL subroutines and GSL thus needs to be installed on your system (download GSL for free from http://www.gnu.org/software/gsl). Before running the program, you also need to set up two input files in the same directory as you put CIFCOX.c. One file, named as *parameter.txt*, sets up the parameters and the other file, *data.txt*, contains the observed competing risks data.

1. Parameter data *parameter.txt*: The file is constructed as follows:

Line	Description	Example
1	Level of partitions in MPT	5
2	Smoothing parameter in MPT	1.0
3		

The first two lines are for the practical setting in MPT. According to Hanson (2006), level in MPT can be approximately equal to $log_2(n/N)$, where n is the sample size of observed data and N is a typical number of observations falling into each partition at the bottommost level, such as 10. Smoothing parameter is considered to be 1, as a sensible canonical choice in Lavine (1992). However, sensitivity analysis should be considered via several different values. Line 3 represents the sample size of your competing risks data. The program also requires the number of covariates in line 4. Line 5 is the total number of MCMC iterations, including the number for burn-in. The updating scheme of all the parameters in this method relies on the Metropolis-Hastings Algorithm (Chib and Greenberg, 1995). The corresponding tuning parameter for each of them needs to be manually adjusted in line 5-9. The acceptance rate should be typically around 20%-40%. The number of acceptances is reported in the output file *accept.txt* (see below). However, Hanson (2006) recommended the acceptance rate for updating Polya trees could be about 40% to 60%

Time	Covariate 1	Covariate p	Cause
1.5239	0.1339	-0.0881	2
1.1686	0.8644	-1.2870	0
0.4540	-2.3967	-0.6793	1

Output File Format

Output files will be sent to a directory called *output*. Users need to create such a subdirectory under the directory containing the CIFCOX.c and the input files. The *output* directory has the acceptance file (*accept.txt*), the files containing the samples from MCMC chains (*coef1.txt*, *coef2.txt*, *mu.txt* and *p.txt*) and the file containing the predictive cumulative incidence function(*predCIF.txt)*.

1. *accept.txt*: The file contains the numbers of acceptance for all the updated parameters. The acceptance rates can be calculated via such numbers divided by the number of MCMC iterations. The first part is the numbers for updating the Polya trees, from the partitions at the bottommost level to ones at the uppermost level and from right to left at each level. The total number of partitions is $2^{M+1} - 2$, where M is the level specification. Since the updates are only required for the partitions with odd numbers, the numbers of acceptances are applied to these odd numbers. The columns next to the label are the acceptance numbers for cause 1 and 2, respectively:

Label	Cause 1	Cause 2
Polya trees 1	5218	4834
Polya trees 3	5398	3742
Polya trees 5	5896	4018

   The next lines are the numbers for updating the normalizing constant (p), parameters (*mu*) in centering distribution for cause 1 and cause 2, coefficients for cause 1 (*coef1*) and coefficients for cause 2 (*coef2*).

2. *coef1.txt*: The file contains p columns of samples over the MCMC iterations for cause of interest.

3. *coef2.txt*: The file contains the coefficient samples for the secondary cause and also has *p* columns.

4. *mu.txt*: The file contains the samples of parameters in the centering distributions. The first column is for the mean parameters of exponential distributions for cause 1 and the second column is for the ones for cause 2.

5. *p.txt*: The file contains the samples of the normalizing constant.

6. *predCIF.txt*: The file contains the predicted baseline cumulative incidence function for cause 1. Since 100 grid points are used in the calculation for each iteration, the file has 100 columns. At each grid point, the mean of the iterations after a burn-in can be treated as the predicted cumulative probability and 2.5th percentile to 97.5th percentile as pointwise 95% credible interval. One can also compute a simultaneous confidence band from the posterior samples.

References

Chib, S. and Greenberg E. (1995). Understanding the Metropolis-hastings Algorithm. *The American Statistician* 49, 327-335.

Fan, X. (2008). Bayesian Nonparametric Inference for Competing Risks Data. Ph.D. Thesis, Medical College of Wisconsin, Milwaukee.

Hanson, T. (2006). Inference for Mixtures of Finite Polya Tree Models. *Journal of the American Statistical Association* 101, 1548-1565.

Lavine, M. (1992). Some Aspects of Polya Tree Distributions for Statistical Modeling. *The Annals of Statistics* 20, 1222-1235.