

Documentation for TWOLEVEL.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 two-level hierarchical model for the Cox-type regression on cumulative incidence function under the competing risks setting. Methods, described in Section 5.2.2 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 TWOLEVEL.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	Sample size for competing risks data	270
4	Number of covariates at individual level	2
5	Number of covariates at center level	1
6	Number of centers	9
7	Number of MCMC iterations	10000
8	Tuning parameters for sampling Polya Trees	0.2 0.2
9	Tuning parameters for centering distributions in MPT	0.05 0.04
10	Tuning parameters for updating coefficients at individual level of cause 1	0.04 0.01
11	Tuning parameters for updating coefficients at individual level of cause 2	0.3 0.3
12	Tuning parameters for updating coefficients at center level of cause 1	0.3

Line	Description	Example
13	Tuning parameters for updating coefficients at center level of cause 2	0.4
14	Tuning parameters for random effects	0.09 0.1
15	Upperbounds of uniform for std. of random effects	2.3 1.8
16	Parameters of beta distribution for updating normalized constant	3.0 7.0
17	Initial values for parameters of centering distribution	1.0 1.0

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 to give the number of covariates at each level in lines 4 and 5. Line 6 require the number of center size. Line 7 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 8-16. 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% and may increase as the level of partitions increases. In this program, MPT priors are assigned on the normalized baseline cumulative incidence functions. The centering distributions of MPT priors are chosen to be exponential. Line 9 represents the tuning parameters for updating the mean of exponential distributions. Lines 10-13 are the tuning parameters for coefficients at the individual and the center level for both causes. For random effects, the updating uses a block scheme, that is, all the random effects are updated at a time for each cause. Correspondingly, the values of tuning parameters are required in line 14. Line 15 represents the values of the upperbound of the prior uniform distributions on the standard deviations of random effects for both causes. Line 16 denotes the two parameters of the beta prior for the normalizing constant. Mean of the beta distribution can be initially set as the percentage of failures due to cause of interest among the exact (not censored) observations. Line 17 represents a reasonable initial guess for the mean parameter of the exponential distributions which are the centering distributions of MPT in this implementation.

2. **Competing risks data** *data.txt*: Each row contains failure time, covariates at individual level, covariates at center level, failure cause and center indication for each individual. The dimension for covariates at the individual level, p , should be the same as defined in line 4 of *parameter.txt* and the dimension for covariates at the center level, c , is the same as defined in line 5. Under the competing risks setting, the cause of interest is coded as 1 and failure due to other causes as 2. In the presence of right censoring, the failure cause is coded as 0. The center indication for the last observation has the same value as declared in line 6. For example,

Time	Cov. 1	...	Cov. p	Cov. 1	...	Cov. c	Cause	Center
1.5239	0.1339	...	-0.0881	1	...	-1.6449	2	1
1.1686	0.8644	...	-1.2870	1	...	-1.6449	0	1
0.4540	-2.3967	...	-0.6793	1	...	-1.6449	1	1
	.							
	.							
0.0781	-0.3406	...	-0.9469	0	...	-0.1257	2	9

Output File Format

Output files will be sent to the directory called *output*. Users need to create such a subdirectory under the directory containing the *TWOLEVEL.c* and the input files. The *output* directory has the acceptance file (*accept.txt*) and the files containing the samples from MCMC chains (*coef1.txt*, *coef2.txt*, *center1.txt*, *center2.txt*, *sigma.txt*, *mu.txt* and *p.txt*)

1. *accept.txt*: The file contains the numbers of acceptances for all the updated parameters. The acceptance rates can be calculated via the numbers divided by the number of MCMC iterations. The first part is the numbers for updating 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	5538	4633
Polya trees 3	5918	3432
Polya trees 5	5706	4068
	.	
	.	

The next lines are the numbers for updating the normalizing constant (p), parameters (μ) in centering distribution for cause 1 and cause 2, coefficients at individual level for cause 1, coefficients at individual level for cause 2, coefficients at center level for cause 1, coefficients at center level for cause 2, random effects for both causes and standard deviations (σ) of random effects for both causes.

2. *coef1.txt*: The file contains the coefficient samples at individual level for cause of interest. The number of columns should be equal to the number defined in line 4 of *parameter.txt*.
3. *coef2.txt*: The file contains the coefficient samples at individual level for the secondary cause.
4. *center1.txt*: The file contains the coefficient samples at center level over the MCMC iterations for cause of interest. The number of columns should be equal to the number defined in line 5 of *parameter.txt*.
5. *center2.txt*: The file contains the coefficient samples at center level for the secondary cause.
6. *sigma.txt*: The file contains two columns of samples over the MCMC iterations. The first column is the samples for the standard deviation of random effects for cause of interest. Those for the secondary cause are in column 2.
7. *mu.txt*: The file contains the samples of parameters in the centering distributions. The first column is for the mean parameters of exponential distribution for cause 1 and the second column is for the ones for cause 2.
8. *p.txt*: The file contains the samples of the normalizing constant.

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.