Documentation for CIF.c
Author: Xiaolin Fan
Updated: 8/1/2008
Questions or bug reports can be sent to xfan@mcw.edu

Description

This program is for estimation of cumulative incidence functions without covariates under the competing risks setting. Methods, described in Section 3.2 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/). Also, the source code for adaptive rejection Metropolis sampling (ARMS) (Gilks and Wild, 1992; Gilks et al, 1995) is required. The les **arms.c** and **arms.h** can be download from http://www.maths.leeds.ac.uk/ wally.gilks/adaptive.rejection/web page/Welcome.html and must be saved in the same directory as CIF.c.

Before running the program, you need to set up two input les in the same directory as CIF.c. One le, named as *parameter.txt*, sets up the parameters and the other le, *data.txt*, contains the observed competing risks data.

1. **Parameter data** *parameter.txt*: The le 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	200
4	Number of MCMC iterations	10000
5	Tuning parameters for sampling Polya Trees	0.6 0.6
6	Tuning parameters for centering distributions in MPT	0.1 0.1
7	Distance between two predictive points	0.03
8	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 can be chosen to be 1, as a sensible canonical choice in Lavine (1992). However, sensitivity analysis should be considered via several different values. Generally speaking, MPT priors lead to empirical distributions when the smoothing parameter is close to zero and to parametric distribution when it is large. Line 3 is the total number of MCMC iterations, including the number for burn-in. The updating scheme of Polya trees and centering distributions in this method relies on the Metropolis-Hastings Algorithm (Chib and Greenberg, 1995). The corresponding tuning parameter for

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 CIF.c and the input files. The *output* directory has the acceptance file (*accept.txt*), the files containing the posterior samples from MCMC chains (*mu.txt* and *p1.txt*) and the files containing the predictive cumulative incidence functions (*CIF1.txt* and *CIF2.txt*).

1. *accept.txt*: The file contains the numbers of acceptances 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,

5. *CIF2.txt*: The le contains the predicted cumulative probabilities for cause 2. It also has 100 columns. Similar calculations can be made as for *CIF1.txt*.

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.

Gilks, W. R., Best, N. G. and Tan, K. K. C. (1995). Adaptive Rejection Metropolis Sampling within Gibbs Sampling. *The Annals of Statistics* 44, 455-472.

Gilks, W. R. and Wild, P. (1992). Adaptive Rejection Sampling fof Gibbs Sampling. *The Annals of Statistics* 41, 337-348.

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.