Documentation for CHANGEPOINT.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 with a time change-point, in the context of violation of the proportionality assumption under the competing risks setting. Methods, described in Section 5.3.1 of Fan (2008), use the Mixture of Polya Trees (MPT) process prior 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 need to set up two input files in the same directory as CHANGEPOINT.c. One file, named as *parameter.txt*

Time	Covariate	Cause
0.4152	0.0	1
2.2329	1.0	0
0.7134	0.0	2

Note that in the case of discrete covariate, 0 stands for the baseline value.

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 CHANGEPOINT.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*), the files containing the predictive distributions (*pred0.txt* and *pred1.txt*) and the file containing LPML values (*LPML.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	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 the centering distributions for cause 1 and cause 2, coefficients for cause 1 (*coef1*) and coefficient for cause 2 (*coef2*).

2. *coef1.txt*: The le contains two columns of samples over the MCMC iterations for cause of interest. The rst column is the samples for the coe cient t before the time change-point and the second column is for the coe cient t after the time change-point.

3. *coef2.txt*: The le contains the coe cient samples for the secondary cause.

4. *mu.txt*: The le contains the samples of parameters in the centering distributions. The rst 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 le contains the samples of the normalizing constant.

6. *pred0.txt*: The le contains the predicted baseline cumulative incidence function for cause 1. Since 100 grid points are used in the calculation for each iteration, the le has 100 columns. At each grid point, the mean of the iterations after a burn-in can be treated as the estimated cumulative probability and 2.5th percentile to 97.5th percentile as the pointwise 95% credible interval. One can also compute a simultaneous con dence band from the posterior samples.

7. *pred1.txt*: The le contains the predicted cumulative incidence function for cause 1 when the value of the discrete covariate is 1. It also has 100 columns. Similar calculations can be made as for *pred0.txt*.

8. *LPML.txt*: The le contains the value of LPML calculated at each iteration. The optimal change-point can be found based on the LPMLs by tting a set of change-