Analysis Example and R Codes

This tutorial provides brief information on the implementation of the estimation procedure for the Bayesian multilevel joint model (BMT-JM) using a sample dataset with 200 facilities. The sample dataset (`sample_data.rdata`) and the following R codes for fitting BMT-MJM are publicly available at https://github.com/esrakurum/BMT-JM:

- `prepare_data.R` – prepares the data to be fit via the joint models,
- `jags.R` – includes the R code for joint modeling via JAGS,
- `Fit.R` – includes the main code; i.e., loading packages and the data, running the code to prepare the data, running the joint model in JAGS, and saving the posterior samples.

In order to run the Bayesian multilevel joint model (BMT-JM),

1. Download all the files and place them in one folder,
2. Set this folder as the working directory using the `setwd(...)` specified in the `Fit.R`,
3. Run the code in `Fit.R`.

The following function within `Fit.R` file fits the BMT-JM models using the Bayesian estimation method presented in the main paper.

```
BMTJM_fit(data, n.iter, n.burn, n.thin, n.chains)
```

The input argument `data` is a data frame containing the variables for the model. In the `BMTJM_fit()` function, `n.iter`, `n.burn`, and `n.thin` is the total number of iterations (including burn-in period), number of iterations during burn-in, and number of iterations for thinning, respectively. The function returns the posterior samples, stored in `post.samples` data frame, obtained from a number of chains indicated using the `n.chains` arguments. In the `post.samples`,

the parameters `Bs.betaX0, Bs.betaX1, Bs.betaX2, Bs.betaZ1, Bs.betaZ2, Bs.h0, Bs.gammaX1, Bs.gammaX2, Bs.gammaZ1, Bs.gammaZ2, Bs.alpha` are the posterior samples for the B-spline coefficients, and `sigmasq_S` and `sigmasq_F` are the posterior samples for σ_S^2 and σ_F^2, respectively. Note that the notation used to store the parameters follows the notation in the main manuscript.

The structure of the input `data` is illustrated in the `sample_data.rdata`:

```
> load("sample_data.rdata")
> head(data)
> names(data)
[1] "subject_id" "facility_id" "time" "x1" "x2" "z1" "z2" "y" "survival" "time_surv"
```

Subject and facility ids in this data set are stored under names `subject_id` and `facility_id`, respectively. The longitudinal measurement times and event times are recorded under `time` and `time_surv`, respectively. The notations for subject-and facility-level covariates follow the main manuscript, that is, `x` and `z`. The event indicator is stored under the column `survival`.