simrec: An R-Package for Simulation of Recurrent Event Data

Katharina Ingel  
IMBEI Mainz

Stella Preussler  
IMBEI Mainz

Antje Jahn-Eimermacher  
IMBEI Mainz

Abstract

**simrec** allows simulation of recurrent event data following the multiplicative intensity model described in Andersen and Gill [1] with the baseline hazard being a function of the total/calendar time. To induce between-subject-heterogeneity a random effect covariate (frailty term) can be incorporated.

**Keywords:** recurrent event data, simulation, total-time model.

Description

This function allows simulation of recurrent event data following the multiplicative intensity model described in Andersen and Gill [1] with the baseline hazard being a function of the total/calendar time. To induce between-subject-heterogeneity a random effect covariate (frailty term) can be incorporated. Data for individual $i$ are generated according to the intensity process

$$Y_i(t) \cdot \lambda_0(t) \cdot Z_i \cdot \exp(\beta'X_i),$$

where $X_i$ defines the covariate vector and $\beta$ the regression coefficient vector. $\lambda_0(t)$ denotes the baseline hazard, being a function of the total/calendar time $t$, and $Y_i(t)$ the predictable process that equals one as long as individual $i$ is under observation and at risk for experiencing events. $Z_i$ denotes the frailty variable with $(Z_i)_i$ iid with $E(Z_i) = 1$ and $Var(Z_i) = \theta$. The parameter $\theta$ describes the degree of between-subject-heterogeneity. Data output is in the counting process format.

**Parameters**

- $N$ Number of individuals
- $fu.min$ Minimum length of follow-up.
- $fu.max$ Maximum length of follow-up. Individuals length of follow-up is generated from a uniform distribution on $[fu.min, fu.max]$. If $fu.min=fu.max$, then all individuals have a common follow-up.
- $cens.prob$ Gives the probability of being censored due to loss to follow-up before $fu.max$. For a random set of individuals defined by a $B(N, cens.prob)$-distribution,
the time to censoring is generated from a uniform distribution on \([0, \text{fu.max}]\). Default is \(\text{cens.prob}=0\), i.e. no censoring due to loss to follow-up.

- **dist.x** Distribution of the covariate(s) \(X\). If there is more than one covariate, \(\text{dist.x}\) must be a vector of distributions with one entry for each covariate. Possible values are "binomial" and "normal", default is \(\text{dist.x}="\text{binomial}"\).

- **par.x** Parameters of the covariate distribution(s). For "binomial", \(\text{par.x}=c(\mu, \sigma)\) where \(\mu\) is the mean and \(\sigma\) is the standard deviation of a normal distribution. If one of the covariates is defined to be normally distributed, \(\text{par.x}\) must be a list, e.g. \(\text{dist.x}<-c(\text{"binomial"}, \text{"normal"})\) and \(\text{par.x}<-\text{list}(0.5, c(1,2))\). Default is \(\text{par.x}=0\), i.e. \(x=0\) for all individuals.

- **beta** Regression coefficient(s) for the covariate(s) \(x\). If there is more than one covariate, \(\text{beta}\) must be a vector of coefficients with one entry for each covariate. \(\text{simrec}\) generates as many covariates as there are entries in \(\text{beta}\). Default is \(\text{beta}=0\), corresponding to no effect of the covariate \(x\).

- **dist.z** Distribution of the frailty variable \(Z\) with \(E(Z) = 1\) and \(\text{Var}(Z) = \theta\). Possible values are "gamma" for a Gamma distributed frailty and "lognormal" for a lognormal distributed frailty. Default is \(\text{dist.z}="\text{gamma}"\).

- **par.z** Parameter \(\theta\) for the frailty distribution: this parameter gives the variance of the frailty variable \(Z\). Default is \(\text{par.z}=0\), which causes \(Z \equiv 1\), i.e. no frailty effect.

- **dist.rec** Form of the baseline hazard function. Possible values are "weibull" or "gompertz" or "lognormal".

- **par.rec** Parameters for the distribution of the event data. If \(\text{dist.rec}="\text{weibull}"\) the hazard function is
  \[
  \lambda_0(t) = \lambda \cdot \nu \cdot t^{\nu-1},
  \]
  where \(\lambda > 0\) is the scale and \(\nu > 0\) is the shape parameter. Then \(\text{par.rec}=c(\lambda, \nu)\). A special case of this is the exponential distribution for \(\nu = 1\).
  If \(\text{dist.rec}="\text{gompertz}"\), the hazard function is
  \[
  \lambda_0(t) = \lambda \cdot \exp(\alpha t),
  \]
  where \(\lambda > 0\) is the scale and \(\alpha \in (-\infty, +\infty)\) is the shape parameter. Then \(\text{par.rec}=c(\lambda, \alpha)\). If \(\text{dist.rec}="\text{lognormal}"\), the hazard function is
  \[
  \lambda_0(t) = \frac{1}{\sigma t} \cdot \frac{\phi\left(\frac{\ln(t)-\mu}{\sigma}\right)}{\Phi\left(-\frac{\ln(t)-\mu}{\sigma}\right)},
  \]
  where \(\phi\) is the probability density function and \(\Phi\) is the cumulative distribution function of the standard normal distribution, \(\mu \in (-\infty, +\infty)\) is a location parameter and \(\sigma > 0\) is a shape parameter. Then \(\text{par.rec}=c(\mu, \sigma)\). Please note, that specifying \(\text{dist.rec}="\text{lognormal}"\) together with some covariates does not specify the usual lognormal model (with covariates specified as effects on the parameters of the lognormal distribution resulting in non-proportional hazards), but only defines the baseline hazard and incorporates covariate effects using the proportional hazard assumption.
• \texttt{pfree} Probability that after experiencing an event the individual is not at risk for experiencing further events for a length of \texttt{dfree} time units. Default is \texttt{pfree}=0.

• \texttt{dfree} Length of the risk-free interval. Must be in the same time unit as \texttt{fu.max}. Default is \texttt{dfree}=0, i.e. the individual is continuously at risk for experiencing events until end of follow-up.

Output
The output is a \texttt{data.frame} consisting of the columns:

• \texttt{id} An integer number for identification of each individual

• \texttt{x} or \texttt{x.V1, x.V2, ...} - depending on the covariate matrix. Contains the randomly generated value of the covariate(s) \(X\) for each individual.

• \texttt{z} Contains the randomly generated value of the frailty variable \(Z\) for each individual.

• \texttt{start} The start of interval \([\texttt{start}, \texttt{stop}]\), when the individual starts to be at risk for a next event.

• \texttt{stop} The time of an event or censoring, i.e. the end of interval \([\texttt{start}, \texttt{stop}]\).

• \texttt{status} An indicator of whether an event occurred at time \texttt{stop} (\texttt{status}=1) or the individual is censored at time \texttt{stop} (\texttt{status}=0).

• \texttt{fu} Length of follow-up period \([0, \texttt{fu}]\) for each individual.

For each individual there are as many lines as it experiences events, plus one line if being censored. The data format corresponds to the counting process format.

Details
Data are simulated by extending the methods proposed by Bender et al [2] to the multiplicative intensity model.

Example

```r
R> library(simrec)
R> ### Example:
R> ### A sample of 10 individuals
R> N <- 10
R>
R> dist.x <- c(“binomial”, “normal”)
R> par.x <- list(0.5, c(0,1))
```
### simrec: Simulation of Recurrent Event Data

```r
beta <- c(0.3, 0.2)

### a gamma distributed frailty variable with variance 0.25

dist.z <- "gamma"
par.z <- 0.25

### and a Weibull-shaped baseline hazard with shape parameter lambda=1
### and scale parameter nu=2.

dist.rec <- "weibull"
par.rec <- c(1,2)

### Subjects are to be followed for two years with 20\% of the subjects
### being censored according to a uniformly distributed censoring time
### within \([0,2]\) (in years).

fu.min <- 2
fu.max <- 2
cens.prob <- 0.2

### After each event a subject is not at risk for experiencing further events
### for a period of 30 days with a probability of 50\%.

dfree <- 30/365
pfree <- 0.5

simdata <- simrec(N, fu.min, fu.max, cens.prob, dist.x, par.x, beta,
                  dist.z, par.z, dist.rec, par.rec, pfree, dfree)
print(simdata[1:10,])
```

### References

1. Andersen P, Gill R (1982): Cox’s regression model for counting processes: a large sample study. The Annals of Statistics 10:1100-1120

2. Bender R, Augustin T, Blettner M (2005): Generating survival times to simulate Cox proportional hazards models. Statistics in Medicine 24:1713-1723
Affiliation:
Katharina Ingel
Institute of Medical Biostatistics, Epidemiology and Informatics (IMBEI)
University Medical Center of the Johannes Gutenberg-University Mainz
55101 Mainz, Germany
E-mail: ingel@uni-mainz.de
URL: http://www.unimedizin-mainz.de/imbei