

Simulation of Patient Enrollment and Survival Data

Yuan Zhong

2024-09-11

Introduction

We offer various examples of data generation using the `Simulation_Enroll` function. Survival times are generated from an exponential distribution characterized by a hazard rate λ , and the occurrence of events such as death, progression, and relapse is modelled using an independent binomial distribution with the rate event.

Patient enrollment is structured across multiple groups. Users of the function must define the duration of patient accrual, follow-up periods, and the maximum duration of the trial. The function allows for the specification of the number of groups, and the distribution of patients across these groups can be either uniform or varied. Below, we illustrate different scenarios utilizing this simulation function.

Simulation Examples

Equal numbers of patients enroll

Consider a clinical trial structured to enroll 100 patients over a three-year period. Details of the patient enrollment are outlined below:

- Total number of patients: 100
- Accrual period: 3 years
- Number of groups: 5
- Length of follow-up: 2 years
- Trial duration: 5 years
- Patients per group: 20

Within each group, the simulation parameters are set as follows:

- `lambda = 0.03`: Survival times are simulated from an exponential distribution using `rexp(n = 20, rate = 0.03)`.
- `event = 0.1`: Event occurrences are determined for each patient based on a Bernoulli distribution `rbinom(n = 20, size = 1, prob = 0.1)`.
- `sensor = 0.9`: It is assumed that the survival times for 90% of patients without an event are shorter than the maximum observed survival time.

The enrollment times for patients in each group are generated using a uniform distribution that is between from the start of the enrollment period to the start of the next group's enrollment. For instance, the enrollment times for the first group are simulated using `runif(n = 20, min = 0, max = 0.6)`. Since the start times across groups are different and sequential, the maximum survival times are adjusted accordingly; the first group's survival times are capped at 5 years, while the second group's are capped at 4.4 years.

```
data <- Simulate_Enroll(n = 100,  
                        lambda = 0.03,  
                        event = 0.1,  
                        M = 1,  
                        group = 5,
```

```

maxt = 5,
accrual = 3,
censor = 0.9,
followup = 2,
partition = "Even")

```

```
head(data)
```

```

##   Time Censor   Enroll
## 1    5      1 0.34977115
## 2    5      1 0.12459001
## 3    5      0 0.25840038
## 4    5      0 0.05199108
## 5    5      0 0.06632739
## 6    5      0 0.10330928

```

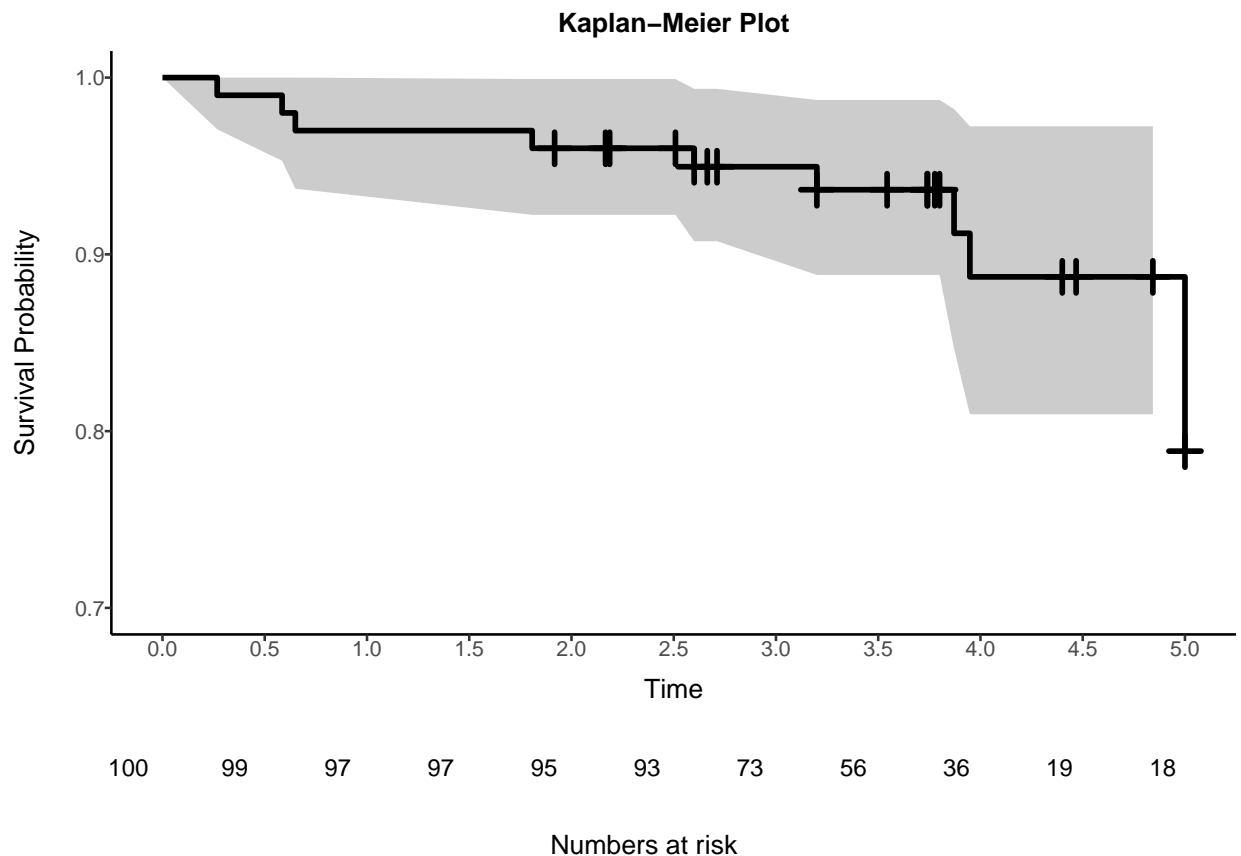
The first column Time indicates the survival times, which are truncated at their respective maximum values. The second column indicates the occurrence of events, with 1 representing an event occurrence and 0 indicating none. The final column records the enrollment time points for each patient. The data in the first two columns are utilized for survival analysis, while the last column is intended for conducting group sequential interim analysis. The KM plot can be shown as Figure 1.

Figure 1. The Kaplan Meier curve of simulated survival data.

```

library(reportRmd)
ggkmCIF(response = c('Time','Censor'), ylim = c(0.7, 1), data=data,
conf.curves = TRUE)

```



Unequal numbers of patients enroll

In many trials, the patient enrollment can have different numbers of patients at different time points. The function can equally partition the enrollment period into 6 time intervals and allow different numbers of patients to enroll in the trial.

```
data <- Simulate_Enroll(n = c(30,20,20,15,10,5),
                      lambda = 0.05,
                      event = 0.2,
                      M = 1,
                      group = 6,
                      maxt = 4,
                      accrual = 3,
                      censor = 0.9,
                      followup = 1,
                      partition = "Uneven")

head(data)
```

```
##      Time Censor   Enroll
## 1 0.2259388     1 0.1548729
## 2 0.4356441     1 0.1081062
## 3 1.2260653     1 0.3104765
## 4 1.4647987     1 0.3402409
## 5 1.5719215     1 0.1234331
## 6 2.3642765     1 0.1832098
```

Consider a clinical trial recruiting different numbers of patients in different time intervals over a three-year period. Details of the patient enrollment are outlined below:

- Total number of patients: 100
- Accrual period: 3 years
- Number of groups: 5
- Length of follow-up: 1 years
- Trial duration: 4 years
- Number of patients in each group: 30,20,20,15,10, and 5 patients

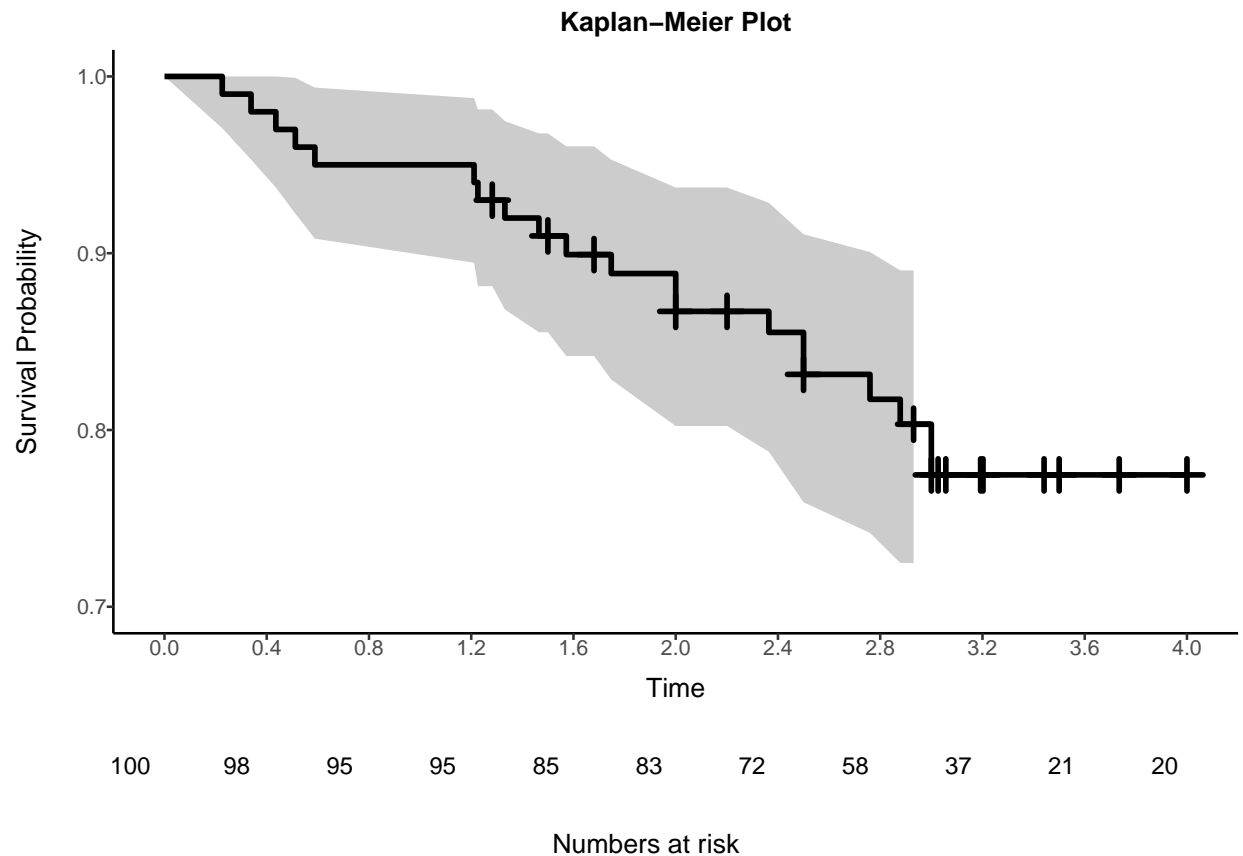
Within each group, the simulation parameters are set as follows:

- `lambda = 0.05`: Survival times are simulated from an exponential distribution using `rexp(n = 20, rate = 0.05)`.
- `event = 0.2`: Event occurrences are determined for each patient based on a Bernoulli distribution `rbinom(n = 20, size = 1, prob = 0.2)`.
- `censor = 0.9`: It is assumed that the survival times for 90% of patients without an event are shorter than the maximum observed survival time.

The enrollment time points and maximum values in each group are generated in the same process as the example above. The KM plot can be shown as Figure 2.

Figure 2. The Kaplan Meier curve of simulated survival data.

```
ggkmCIF(response = c('Time','Censor'), ylim = c(0.7, 1), data=data,
        conf.curves = TRUE)
```



Multiple streams of data generating

The function can generate multiple survival outcomes based on the same parameter settings. M stands for the number of streams for MCMC. For example, we can set $M = 4$ to generate four independent simulation outputs.

```
data <- Simulate_Enroll(n = c(30,20,20,15,10,5),
  lambda = 0.05,
  event = 0.2,
  M = 4,
  group = 6,
  maxt = 4,
  accrual = 3,
  censor = 0.9,
  followup = 1,
  partition = "Uneven")
head(data[[1]])
```

```
##      Time  Censor   Enroll
## 1 0.5664612      1 0.38337001
## 2 0.9889685      1 0.21845758
## 3 1.4329438      1 0.26098106
## 4 1.5122553      1 0.01593174
## 5 2.3543654      1 0.15919320
## 6 2.5314415      1 0.01324882
```

```
head(data[[2]])
```

```
##           Time  Censor      Enroll
## 1 0.3152769      1 0.30699998
## 2 3.1449971      1 0.03929211
## 3 3.6528345      1 0.41792985
## 4 3.6713184      1 0.06701538
## 5 3.9342905      1 0.30207896
## 6 4.0000000      1 0.31204986
```

```
head(data[[3]])
```

```
##           Time  Censor      Enroll
## 1 0.2050598      1 0.20269500
## 2 2.4995733      1 0.06881723
## 3 3.7622478      1 0.36650349
## 4 4.0000000      1 0.48914029
## 5 4.0000000      1 0.01495786
## 6 4.0000000      1 0.06608268
```

```
head(data[[4]])
```

```
##           Time  Censor      Enroll
## 1 0.4949577      1 0.44314315
## 2 0.9691593      1 0.43092560
## 3 1.0979019      1 0.08215292
## 4 1.6856785      1 0.47052760
## 5 2.6682526      1 0.41463774
## 6 3.5676237      1 0.38016145
```