The Kaplan–Meier theatre

Thomas A. Gerds

Department of Biostatistics, University of Copenhagen, Copenhagen, Denmark.
e-mail: tag@biostat.ku.dk

Summary Survival is difficult to estimate when observation periods of individuals differ in length. Students imagine sailing the Titanic and then recording whether they "live" or "die". A clever algorithm is performed which results in the Kaplan-Meier estimate of survival.

Keywords: Teaching; Censored data; Medical statistics; Kaplan–Meier method; Survival probability function.

INTRODUCTION

Survival probabilities are not straightforward to obtain when observation periods of individuals differ in length. The Kaplan–Meier theatre is a classroom activity, which starts by a data collection exercise where students imagine sailing on the Titanic. Several students 'fall in the water' where they are observed by a neighbouring student while they try to hold their breath as long as they can. The observation periods are designed such that some students 'drown' and other 'survive' until the end of the experiment. Based on the data collected, it is explained why even simple statistics may fail when applied naively. For example, the frequency of students who 'survived' 40 s would generally be an estimate of the probability to survive 40 s. However, an issue occurs when there is a student who 'survived' but was observed only for 35 s. Then, it is unknown (censored) if the student 'drowned' between 35 and 40 s. The Kaplan–Meier method assumes that censored individuals have the same survival chances as the individuals who are still observed. During the Kaplan–Meier theatre, students perform a clever algorithm (Efron 1967), which translates the assumption into action and results in the Kaplan–Meier estimate of survival.

The activity requires little preparation from the teacher and no special equipment (Appendix).

ON THE TITANIC

In the following, I am the teacher, and the text in *italics* refers to what I would say during the class. The original form of the story to be told is the following.

We are all on the Titanic, and the Titanic is going down. Once under water, you would have to hold your breath. But how long can you do this? Every second person in the room will participate as an actor in the theater. The nearest non-participating neighbor will act as a timekeeper and count the number of seconds the participants can hold their breath. Every timekeeper needs a mobile phone or another time-tracking device. Nonetheless we will not all start simultaneously. Because the Titanic is sinking slowly, the participants touch the water at different points in time. As soon as each participant has found a timekeeper, I will start the experiment. When I knock on the table of the first participant, the experiment has begun. The first participant starts holding her breath. About five seconds later I knock on the next table. This is when the second participant is "falling in the water" and starts holding his breath. Thus, the gap between each pair of adjacent participants is roughly 5 seconds. I move through the classroom in this way until all participants have started to hold their breath. The participants continue to try to hold their breath until I say "stop". This is the end of the follow-up period of the experiment. Thus the individual follow-up period stops for all participants at the same point in time. We can imagine that the person who was recording the data on the Titanic had to save herself at some critical time, and thus only data are available up to that time point.

We then collect the data. Participants who were not able to hold their breath until the stop-time *are said to have "drowned". For those who have "drowned" we will need the number of seconds between "falling in the water" and the moment when they no longer could hold their breath. Participants who were able to hold their breath until the stop-time are said to have "survived". For those who "survived" we will need the number of seconds until the end of the follow-up period. Once everyone is set and ready, we start to let the Titanic sink.*

The Kaplan–Meier theatre has been done with 8 to 12 participants, and it has worked well for this size. The goal is to have a mixture of participants who experience an event within the study period ('drowned') and participants who do not ('survived'). In a given class of students, the outcome will depend on the athletic condition and personal ambition of the students and is generally hard to forecast. However, the results are influenced by the design of the experiment. Useful results are expected, and were achieved in previous performances of the activity, when the follow-up period is stopped about 80–90 s after the first participant was 'falling in the water', that is about 30–40 s after the last participant was 'falling in the water'. To be sure that there are at least some drowning events, one could arrange with the participants beforehand that they should raise their hand as soon as they can no longer hold their breath.

WHEN SIMPLE SUMMARY STATISTICS FAIL

Participants who "drowned" write the number of seconds with at most one decimal place in large red colored letters on a large (letter size/din A4) piece of paper. Participants who "survived" write the number of seconds in large green colored letters. All participants take their piece of paper and come to the front.

Ideally, the participants can be lined up in front of the rest of the class, such that the teacher can still write on a broadly visible part of the board.

The first task is to sort the line of participants from left to right in increasing order of the number of seconds written on their papers.

This action is sketched in figure 1, which shows a specific example of the results of the Kaplan–Meier theatre. The data shown will be used for illustration in what follows. If a time to drowning and time to not-drowning tie, we put the drowning time first. That is why in the lower panel of figure 1, the person with the red 42 s is standing to the left of the person with the green 42 s. Once the participants are sorted, the first question goes to the people who are still sitting (timekeepers and other non-participants). Here are proposals for the first question:

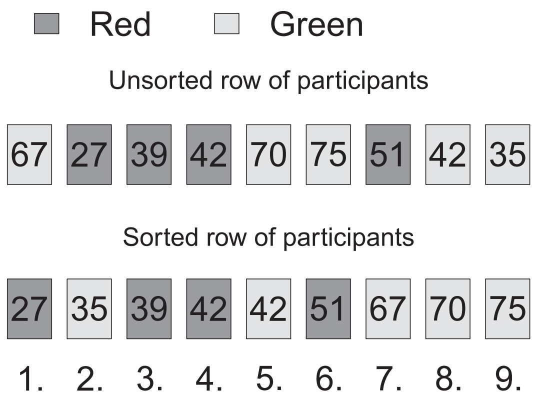

Fig. 1. Example of the results of a Kaplan–Meier theatre. The coloured boxes represent the pieces of paper with the number of seconds written in either light-grey/green ('survived' until end of follow-up) or dark-grey/red ('drowned'). The upper panel illustrates the situation when the nine participants have lined up in front of the board. The lower panel shows the same participants when they have sorted according to the number of seconds

Based on the data that you see,

- *what is the probability of surviving 40 seconds?*
- *what is the median survival time?*
- *what is the mean survival time?*

The aim of the first question is to provoke a wrong answer. Thus, in order to increase the likelihood of receiving a wrong answer to the first of the proposed questions there should be at least one of the participants, a survivor, who is lost to follow-up before 40 s. The value 40 is chosen for illustration of the example presented here. The first question can be phrased with any value, which is both larger than the smallest green number and smaller than the largest number in the data set. The example data shown in figure 1 satisfy this requirement because one of the participants has written 35 s in green on the paper he is holding. Note that the answer to the question regarding the mean survival time is wrong as soon as there is at least one green number. If really no one is willing to give a wrong answer, one can proceed anyway by asking:

Why would it be wrong to estimate the probability to survive 40 seconds by the relative frequency of participants who have more than 40 seconds on their piece of paper?

After some reflection, someone in the class will discover the problem, namely, that for the participants who have a green number of seconds, which is smaller than 40, it is unknown when the event happened. Everyone should now observe that the difference between the colours. Red numbers represent observed event times, whereas green numbers represent right censored observations for which the event time is unknown. Right censored means that the only information available is that the event time is greater (to the right on the time scale) than the green number. The teacher will then introduce that a clever algorithm (the Kaplan–Meier method) solves this problem.

THE REDISTRIBUTION ALGORITHM IN ACTION

A survival function is a function that for a given time point returns a survival probability. The Kaplan–Meier estimate of the survival function will be a jump function with (downward) jumps only at time points were at least one 'drowning event' was observed. In what follows, the symbol C is used to mark censored observations. Initially, assign a probability mass of 1/9 to each of the nine observations:

27	35 (C)	39	42	42 (C)	51	67 (C)	70 (C)	75 (C)
1/9	1/9	1/9	1/9	1/9	1/9	1/9	1/9	1/9

I now instruct the participants to perform the calculation of Efron's redistribution to the right algorithm. To keep track of the results, I first draft an empty table on the board with the following columns: *Time*, *Number of subjects in study*, *Number of events*, *Number lost to follow-up*, *Survival probability (%)*. Then I fill the first row with the information. which is available at time zero. Table 1 shows the final version of the table for the data in figure 1.

Initially each piece of paper weighs one over nine because there are 9 participants. We start with the participant standing most to the left. You have "drowned" after 27 seconds. At the time where you "drowned" all other 8 participants were still alive and still in the study. Thus, at 27 seconds the Kaplan–Meier estimate of survival drops from 100% by 1/9 and takes on the value 88.9%. Please take your piece of paper and sit down.

Then I adress the next participant in line.

You "survived" 35 seconds, and then we lost track of what happened to you. We may assume that you "drowned" at a later time point but we do not know when. The Kaplan–Meier method assumes that your survival chances after 35 seconds are equal to those of the remaining seven participants standing to your right. Thus, you tear your piece of paper into 7 equally large pieces and give one piece to each of the remaining seven participants.

It is important that each of the seven participants receives a piece of paper, however not so important that the pieces are exactly equally large. In order to aid the computations at later stages of the algorithm, the participant can write the fractional weight of the distributed paper pieces on the back side of each of the seven pieces of paper. The distribution of probability mass is now:

27	35 (C)	39	42	42 (C)	51	67 (C)	70 (C)	75 (C)
1/9	0	1/9 + 1/7 1/9	1/9 + 1/7 1/9	1/9 + 1/7 1/9	1/9 + 1/7 1/9	1/9 + 1/7 1/9	1/9 + 1/7 1/9	1/9 + 1/7 1/9

I write line 3 of table 1, and then address the next participant in line.

You have "drowned" after 39 seconds. To account for the possibility that participant 2 may have drowned between 35 seconds and 39 seconds, our estimate of the survival probability now drops by the total weight of the paper in your hand:

$$\underbrace{1/9}_{\text{own contribution}} + \underbrace{1/7\times 1/9}_{\text{contribution from participant no.2}}$$

Table 1. Kaplan–Meier analysis of the data shown in figure 1

Time (seconds)	Number of subjects	Number of events	Number lost to follow-up	Survival probability (%)	Calculation
0	9	0	0	100	everyone alive
27	9	1	0	88.9	1–1/9
35	8	0	1	88.9	no change
39	7	1	0	76.2	0.889–1/9–1/7 1/9
42	6	1	1	63.5	0.762–1/9–1/7 1/9
51	4	1	0	47.6	0.635–1/9–1/7 1/9–1/4 (1/9–1/7 1/9)
67	3	0	1	47.6	no change
70	2	0	1	47.6	no change
75	1	0	1	47.6	no change

The data and treatment of participant 4 leads to another drop of the same amount in survival (Table 1). Participant number 5 is instructed to divide both the owned piece of paper and the one from participant number 2 into four equal pieces and to pass an equal amount of paper to each of the remaining participants. When participant 6 'drowns', the drop in survival is thus according to the weight of all the pieces of paper:

$$\underbrace{1/9}_{\text{own contribution}} + \underbrace{1/7\times 1/9}_{\text{contr. from part.2}} + \underbrace{1/4\times 1/9}_{\text{contr. from part.5}} + \underbrace{1/4\times 1/7\times 1/9}_{\text{contr. from part.5 via part.2}}$$

Participant 7 distributes half of each piece of paper to participants 8 and 9, and participant 8 gives all pieces of papers to participant 9. After 75 s, the Kaplan–Meier estimate is no longer defined. But note that if the last participant would have 'drowned', then the Kaplan–Meier estimate would have dropped all the way down to the value of zero.

THE KAPLAN–MEIER PLOT

Once the theatre drama concludes, I discuss the assumptions of the Kaplan–Meier method (see section on Discussion). Based on the data of table 1, I draw the Kaplan–Meier graph on the board (Figure 2). While doing this, I discuss each step of the line, and mark the line at the time points at which subjects were lost to follow-up (censored). It is useful to pronounce that and repeat why the step size of the Kaplan–Meier graph is increasing over time. Further, one can mention that the definition of the Kaplan–Meier graph stops at the largest observation time and that, depending on the type of the event, it may be more interesting to reverse the y-axis and draw one minus the Kaplan–Meier estimate, i.e. an estimate of the cumulative distribution function. Finally, I ask the students to look at the figure and read off the answers to the introductory questions:

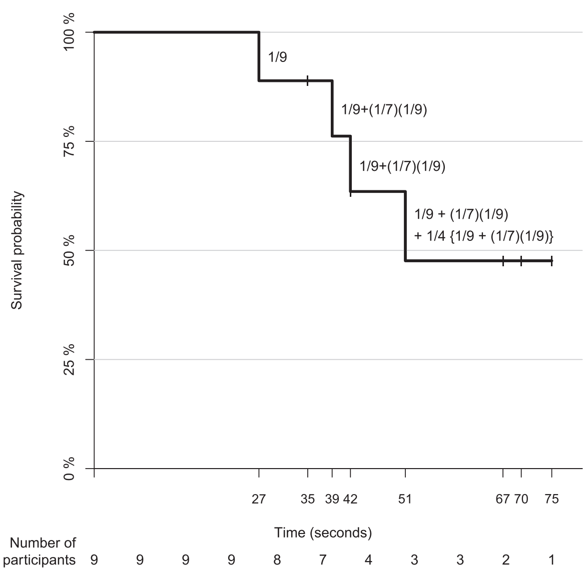

Fig. 2. Kaplan–Meier graph based on the data shown in figure 1

- *What is the probability to survive 40 seconds?* **Answer**: *76.2%.*
- *What is the median survival time?* **Answer**: *51 seconds.*
- *What is the mean survival time?* **Answer**: *not knowable.*

DISCUSSION

In the 1950s, Edward L. Kaplan and Paul Meier derived a solution for estimating survival probabilities from incomplete data (Kaplan and Meier 1958). Today, the method is used in various fields of applied statistics – most prominently in medical and social sciences. For example, the Kaplan–Meier method can be used to describe survival chances of cancer patients.

In fact, the design of the Kaplan–Meier theatre aims to mimic that of a typical medical study. The time at which a participant 'falls into the water' corresponds to the time at which a patient is enrolled into the study. The date of enrollment (start of follow up) is often connected to a medical intervention such as surgery or start of another treatment. The accrual period in which patients are enrolled into a medical study can last for several years. To see the connection with the theatre, suppose that the last patient of a study is enrolled 2 years after the first patient, and the data are collected for statistical analysis 3 years after the first patient was enrolled. Then data are observed within a 3-year follow-up period for the first patient but only within a 1-year follow-up period for the last patient. The survival time is observed for patients who die within the follow-up period but unknown (censored) for patients who are alive at the end of the follow-up period.

During the theatre, students learn that in case of end of follow-up (censored), it makes sense to distribute the probability mass to the right. The reasoning requires that the event ('drowning') really occurs to the censored participant at a later time point. The reasoning would fail if a lifeboat would come to save some participants from 'drowning'. In statistical terminology, an event, which makes a later occurrence of the event of interest impossible, is called a competing risk. In the real world, competing risks are the rule rather than the exception. For example, death from non-cardiovascular causes is a competing risk when the event of interest is cardiovascular disease. Unfortunately, the Kaplan–Meier method cannot handle competing risks. If applied naively, the Kaplan–Meier method would treat a competing risk (death due to other causes/lifeboat) in the same way as it treats end of follow-up (right censored). But, by redistributing probability mass in this case, one would generate a hypothetical world in which the competing risk does not exist. Unfortunately, this mistake has been made many times (e.g. Southern et al. 2006).

When there are competing risks, it is often appropriate to use the Aalen–Johansen method to estimate event probabilities (Aalen and Johansen 1978).

The Kaplan–Meier method estimates the average (marginal) survival probabilities in the population. This is reflected during the theatre play where subjects who were lost to follow up distributed the same amount of the probability mass to all subjects standing to the right, no matter their gender, lung volume and other characteristics, which naturally effect the survival outcome. The method can be modified, for example, by letting women with high lung volumes who are lost to follow up distribute their probability mass only to the remaining women who also have high lung volumes. A corresponding extension of the Efron's redistribution to the right algorithm is described in Malani (1995).

The calculations become extensive quite quickly with larger data sets. Therefore, Appendix provides an *R* function, which writes out the calculations of the redistribution algorithm for a given data set.

References

Aalen, O. and Johansen, S. (1978). An empirical transition matrix for non-homogeneous Markov chains based on censored observations. *Scandinavian Journal of Statistics*, **5**, 141–150.

Efron, B. (1967). The two-sample problem with censored data. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. Berkeley: University of Berkeley Press, **4**, 831–853.

Kaplan, E. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, **53**, 457–481.

Malani, H.M. (1995). A modification of the redistribution to the right algorithm using disease markers. *Biometrika*, **82**, 515–526.

Southern, D.A., Faris, P.D., Brant, R., Galbraith, P. D., Norris, C.M., Knudtson, M.L. and Ghali, W.A. (2006). Kaplan–Meier methods yielded misleading results in competing risk scenarios. *Journal of Clinical Epidemiology*, **59**, 1110–1114.

APPENDIX

Ingredients

The following items are needed to perform the Kaplan-Meier theater:

- A stopwatch (or another time tracking device).
- A bunch of red and green text markers.
- About 10 pieces of paper.
- About 10 mobile phones (or 10 other time tracking devices).
- A white or black board.

R code

The function redist requires the R-library prodlim which can be downloaded from CRAN. Note that the function redist does not require that the data are sorted according to time. In the following example R-code, the values of the status parameter represent "drowned" (1) and "censored" (0) events. It produces all the calculations needed for Table 1.

```
library(prodlim) ## need version >= 1.5.7
time=c(27,35,39,42,42,51,67,70,75)
status=c(1,0,1,1,0,1,0,0,0)
redist(time,status)
plot(prodlim(Hist(time,status)~1))
```