Simfit

Tutorials and worked examples for simulation, curve fitting, statistical analysis, and plotting. http://www.simfit.org.uk

Given observations of some critical event such as survival, recovery from illness, failure of a machine, or death of a subject, as a function of time within a given group, the Kaplan-Meier or product moment nonparametric estimator for a suitable step function to model the survivor function has gained wide acceptance.

Example 1

From the main SIMFIT menu choose [Statistics], [Time series and survival], then [Kaplan-Meier], and study the default test file `survive.tf2` which has the following format.

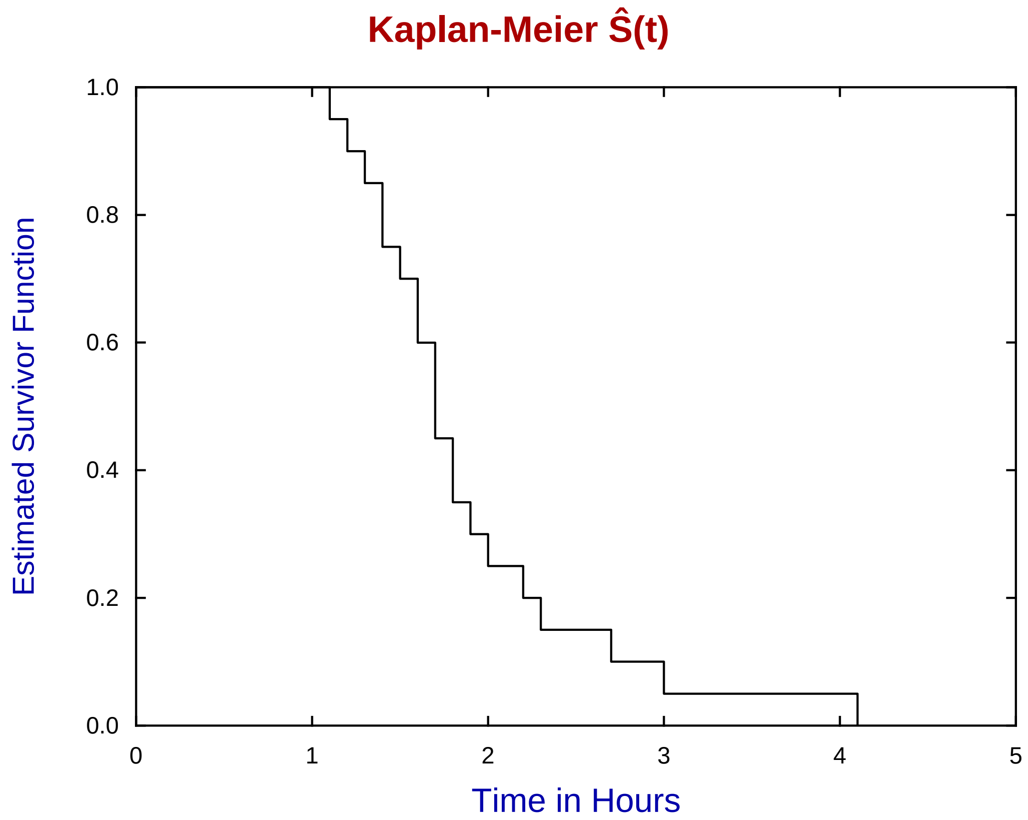
Time to relief	Censorship	Frequency
1.1	0	1
1.4	0	1
1.3	0	1
1.7	0	1
1.9	0	1
1.8	0	1
1.6	0	1
2.2	0	1
1.7	0	1
2.7	0	1
4.1	0	1
1.8	0	1
1.5	0	1
1.2	0	1
1.4	0	1
3.0	0	1
1.7	0	1
2.3	0	1
1.6	0	1
2.0	0	1

These data were for twenty patients taking an analgesic to relieve headache pain and the data have been formatted according to this scheme, where the critical event in this case is freedom from pain.

1. **First column**
   Time in hours (not necessarily ordered)
2. **Second column**
   Censoring code (0 = occurrence of the critical event, 1 = right-censored)
3. **Third column**
   Frequency of the observation
4. **Note**
   The starting sample size will be taken as the sum of all the frequencies in column 3. So subjects remaining at or after the last failure should be included as right-censored with the appropriate frequency. Failure to do this will lead to an underestimate of the starting size for the group and a biased estimator.

These data do not contain censored observations and all the subjects were observed until the headache ceased, so the estimated survivor function $\hat{S}(t)$ was as displayed in the next table and graph.

Time to relief	Estimate $\hat{S}(t)$	Standard Error
1.1	0.95	0.0487
1.2	0.90	0.0671
1.3	0.85	0.0798
1.4	0.75	0.0968
1.5	0.70	0.1025
1.6	0.60	0.1095
1.7	0.45	0.1112
1.8	0.35	0.1067
1.9	0.30	0.1025
2.0	0.25	0.0968
2.2	0.20	0.0894
2.3	0.15	0.0798
2.7	0.10	0.0671
3.0	0.05	0.0487
4.1	0.00	0.0000



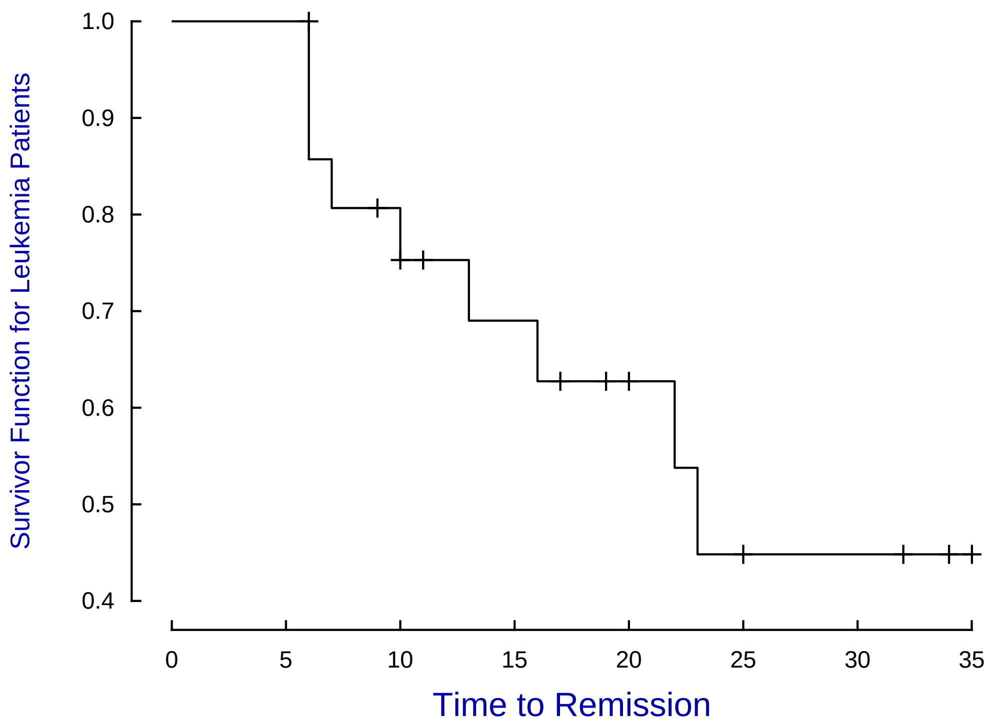
Example 2

The next table and graph are for the data contained in test file `survive.tf1` which is for time to remission in 21 leukemia patients.

This data set contains right-censored observations with a 1 instead of 0 in column two, and also records the number of replicates with frequencies of 2 or 3 instead of 1 in column three.

Time to remission	Censorship	Frequency
6	1	1
6	0	3
7	0	1
9	1	1
10	0	1
10	1	1
11	1	1
13	0	1
16	0	1
17	1	1
19	1	1
20	1	1
22	0	1
23	0	1
25	1	1
32	1	2
34	1	1
35	1	1

Kaplan-Meier $\hat{S}(t)$ [+ indicates censorship]



Note that the eleven points at times 6, 9, 10, 11, 17, 19, 20, 25, 32, 34, and 35 where loss by censoring happened are indicated in the above diagram by plus signs (+).

Theory

The nomenclature regarding the Kaplan-Meier estimator for a survivor function arose because it is most widely used in reliability studies, where machinery is operated until failure occurs, and in clinical studies where a group of patients under treatment is observed until some critical event like recovery from illness, relief from suffering, or death happens. It is often complicated by the fact that right censoring occurs, where a subject leaves the group without the critical event occurring, e.g. when a clinical trial is discontinued.

For these reasons it is well to remember that as $S(t) = 1 - F(t)$ then $F(t) = 1 - S(t)$ and there may be some occasions where it could be more logical to regard $F(t)$ as the *survivor function*.

Suppose that there are exactly k distinct times where at least one critical event, e.g. a failure occurred. Then the calculation is based on ordering these k distinct times for failure into an increasing sequence

$$t_1 < t_2 < t_3 < \cdots < t_k$$

and recording the number that failed at each time, but also taking note of the number lost at each time due to censoring.

To understand the method, note that, as the times t_i are distinct and ordered failure times, i.e. $t_{i-1} < t_i$, and the number in the sample that have not failed by time t_i is n_i, while the number that do fail is d_i, then the estimated probabilities of failure and survival at time t_i are given by

$$\hat{p}(\text{failure}) = d_i/n_i$$
$$\hat{p}(\text{survival}) = (n_i - d_i)/n_i.$$

The Kaplan-Meier product limit nonparametric estimate of the survivor function is then defined as a step function which is given in the interval t_i to t_{i+1} by the product of survival probabilities up to time t_i, that is

$$\hat{S}(t) = \prod_{j=1}^{i} \left(\frac{n_j - d_j}{n_j} \right)$$

with variance estimated by Greenwood's formula as

$$\hat{V}(\hat{S}(t)) = \hat{S}(t)^2 \sum_{j=1}^{i} \frac{d_j}{n_j(n_j - d_j)}.$$

It is understood in this calculation that, if failure and censoring occur at the same time, the failure is regarded as having taken place just before that time and the censoring just after it.

It should be pointed out that steps in the survivor function only occur at failure points and, in the absence of any censored points, the Kaplan-Meier estimate for the survivor function at $t = t_i$ is just the usual binomial parameter estimate.