

# Inference for the Kaplan-Meier Estimator

Patrick Breheny

September 7

# Introduction

- Last time, we discussed the estimation of the survival function  $S(t)$  using a completely nonparametric approach
- As with any estimator, a natural follow-up question is: How reliable is the estimate? (or in other words, how much trust should we have that the truth is close to the estimate?)
- That is our goal for today: to derive standard errors and confidence intervals for  $S(t)$

## Central limit theorem for $\lambda_j$

- We saw last time that the Kaplan-Meier estimate is a function of  $\{\lambda_j\}$  estimates
- Thus, let's begin by deriving the standard error of  $\hat{\lambda}_j$
- Recall that the likelihood for  $\lambda_j$  is simply a binomial likelihood, so

$$\frac{\hat{\lambda}_j - \lambda_j}{\sqrt{\lambda_j(1 - \lambda_j)/n_j}} \xrightarrow{d} N(0, 1)$$

- In other words,  $\hat{\lambda}_j$  is approximately normal with mean  $\lambda_j$  and variance  $\lambda_j(1 - \lambda_j)/n_j$

# Log transformation

- Of course, it isn't immediately clear how all the  $\hat{\lambda}_j$  variances will affect the variance of  $\hat{S}(t)$
- Consider, then, the log of the survival function:

$$\log S(t) = \sum_{t_j \leq t} \log(1 - \lambda_j)$$

- Now the variances will simply add up (provided that the  $\hat{\lambda}_j$ s are independent), although now we need the variance of  $\log(1 - \hat{\lambda}_j)$

## Applying the delta method

- By the delta method,

$$\mathbb{V} \log(1 - \hat{\lambda}_j) = \frac{\mathbb{V}(\hat{\lambda}_j)}{(1 - \lambda_j)^2} = \frac{\lambda_j}{n_j(1 - \lambda_j)}$$

- Therefore,

$$\mathbb{V} \log\{\hat{S}(t)\} = \sum_{t_j \leq t} \frac{\lambda_j}{n_j(1 - \lambda_j)}$$

- In what follows, we will denote this quantity  $\sigma^2(t)$ , with  $\hat{\sigma}^2(t)$  defined similarly, but with  $\hat{\lambda}_j$  replacing  $\lambda_j$   
*Note: the book denotes this quantity  $s(t)$ , which is fine in print, but on the board it is hard to tell  $S(t)$  and  $s(t)$  apart*

## Central limit theorem for $\log S(t)$

- Thus,

$$\frac{\log \hat{S}(t) - \log S(t)}{\sigma(t)} \xrightarrow{d} N(0, 1),$$

where  $\sigma(t)$  is the standard error (defined on the previous slide) at time  $t$

- We are now in a position to construct confidence intervals for  $S(t)$ : the 95% confidence interval for  $S(t)$  is given by

$$\hat{S}(t) \exp\{\pm 1.96 \hat{\sigma}(t)\}$$

## Is this derivation valid?

- The preceding derivation would be perfectly valid if  $T$  followed a discrete distribution with mass points  $t_j$  known in advance
- For a continuous  $T$ , however, not only are the failure times  $t_j$  not known in advance, but as  $n \rightarrow \infty$ , the number of unique failure times increases without bound (as does the number of parameters  $\lambda_j$  that must be estimated)
- Nevertheless, it turns out that the asymptotic normality result we just derived does, in fact, hold for continuous  $T$ , but proving it requires a different version of the central limit theorem called the *martingale central limit theorem* that lies beyond the scope of this course (but see Chapter 5.5-5.6 of Kalbfleish & Prentice for more details)

## Central limit theorem for $S(t)$

- Now that we have a central limit theorem result for  $\log S(t)$ , we can go back (again using the delta method) to obtain a central limit theorem result for  $S(t)$  itself
- Doing so yields

$$\frac{\hat{S}(t) - S(t)}{\sigma(t)S(t)} \xrightarrow{d} N(0, 1)$$

- Thus, we can construct asymptotic 95% confidence intervals for  $S(t)$  using

$$\hat{S}(t) \pm 1.96 \hat{\sigma}(t) \hat{S}(t);$$

this formula for the standard error of  $\hat{S}(t)$  is known as *Greenwood's formula*, and was originally proposed all the way back in 1926



## Criticisms of the Greenwood method

- The Greenwood confidence intervals have a long history and the advantage of simplicity
- An important criticism, however, is that by taking  $\hat{S}(t)$  and adding and subtracting an error term, the bounds of our interval can lie outside  $[0, 1]$ , which is clearly not sensible  $\sigma(t)$
- It is worth pointing out, however, that our first approach doesn't satisfy this criterion either: we can't get negative bounds, but there is nothing preventing  $\hat{S}(t)e^{1.96\hat{\sigma}(t)}$  from being greater than 1

## The complimentary log-log transformation

- Consider, then, one more scale on which we can derive a central limit theorem result
- Consider the transformation

$$g(x) = \log\{-\log(x)\}$$

- This transformation is known as the *complimentary log-log* transformation
- Note that if  $x \in [0, 1]$ , the range of  $g(x)$  is unrestricted; to put it another way, no matter what  $g(x)$  is,  $g^{-1}(x)$  will lie between 0 and 1

# Complimentary log-log CLT

- Once again applying the delta method,

$$\frac{\log\{-\log \hat{S}(t)\} - \log\{-\log S(t)\}}{\sigma(t)/\log S(t)} \xrightarrow{d} N(0, 1)$$

- Thus, by transforming, we can construct 95% confidence intervals for  $S(t)$  via

$$\hat{S}(t)^{\exp\{\pm 1.96 \hat{\sigma}(t) / \log \hat{S}(t)\}}$$

- This isn't the only transformation with this property, of course, but it is the one most widely used in survival analysis

## Examples: Introduction

- Now that we've derived these useful results, let's see how our confidence intervals work in practice
- We'll work out our three confidence interval methods by hand for one time point on the GVHD data, then see what it looks like when the intervals are plotted as bands over top of the KM curve
- Finally, we'll discuss the R code/commands for making this automatic

# GVHD standard errors (MTX+CSP group)

$t$	$n(t)$	$d(t)$	$t$	$\hat{S}(t)$	$\frac{d_j}{n_j(n_j-d_j)}$	$\hat{\sigma}^2(t)$	$\hat{\sigma}(t)$
0	32	0	0	1	0	0	0
8	31	1	8	0.968	$1/(31 \cdot 30)$	0.0011	0.033
10	30	1	10	0.935	$1/(30 \cdot 29)$	0.0022	0.047
16	28	1	16	0.902	$1/(28 \cdot 27)$	0.0035	0.060
...			...				

# Greenwood interval

- Let's start with the simplest method, where we construct CIs directly on the survival probability scale using Greenwood's formula
- At time 9 days,

$$\begin{aligned}\hat{S}(9) \pm 1.96 \hat{\sigma}(9) \hat{S}(9) &= 0.968 \pm 1.96(0.968)(0.033) \\ &= [0.906, 1.030]\end{aligned}$$

- As we remarked earlier, this can happen with Greenwood intervals, where the upper bound is above 1
- The common practice is to clip the interval at  $[0, 1]$  and report  $[0.906, 1]$  as the 95% CI

## Log and log-log intervals

- Let's try the approach where we construct the interval on the log scale and then back-transform:

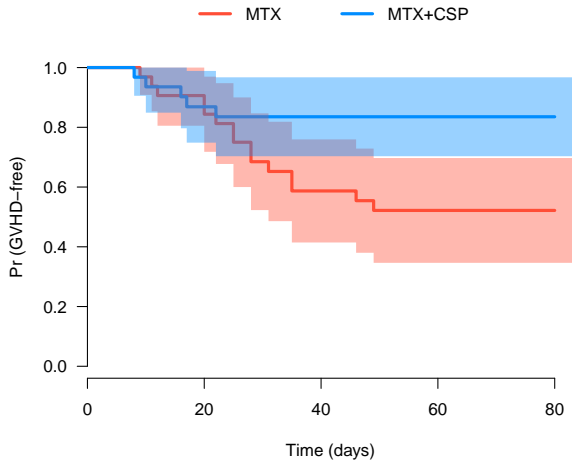
$$\begin{aligned}\hat{S}(9) \exp\{\pm 1.96 \hat{\sigma}(9)\} &= 0.968 \exp\{\pm 1.96(0.033)\} \\ &= [0.908, 1.032]\end{aligned}$$

- Finally, on the complimentary log-log scale:

$$\begin{aligned}\hat{S}(9)^{\exp\{\pm 1.96 \hat{\sigma}(9) / \log \hat{S}(9)\}} &= 0.968^{\exp\{\pm 1.96(0.033) / \log(0.968)\}} \\ &= [0.792, 0.995]\end{aligned}$$

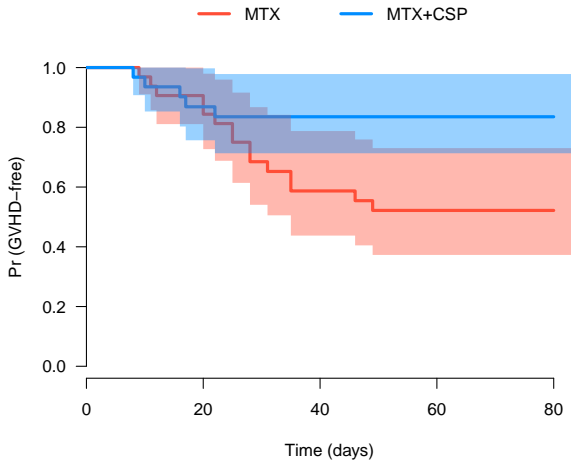
- Note that the complimentary log-log interval is very different than the other two

# GVHD: Greenwood / "plain" / "linear" intervals

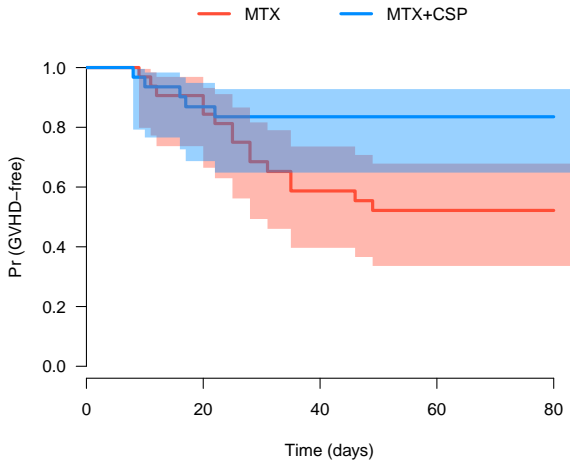




# GVHD: Log intervals



# GVHD: Log-log intervals



## Which interval is best?

- There is no general consensus on which of the three intervals is best; all three are in widespread use and might be the default, depending on the software you're using
- The appeal of the complimentary log-log interval is clear, but the log-scale interval has the advantage of variance stabilization:  $S(t)$  does not appear in the standard error term – for this reason, it is the default in the `survival` package
- Generally, the log and log-log intervals have been shown via simulation to have better finite-sample properties than the Greenwood interval

## CI's with survfit

- Speaking of the `survival` package, let's quickly go through the commands and options for producing CI's for  $S(t)$
- The `survfit` function calculates confidence intervals automatically; you can choose your preferred method using the `conf.type` argument:

```
fit <- survfit(S ~ Group, Data, conf.type='plain')  
fit <- survfit(S ~ Group, Data, conf.type='log')  
fit <- survfit(S ~ Group, Data, conf.type='log-log')
```

where `conf.type='log'` is the default

## Displaying CIs

- Once the curve and its confidence intervals have been estimated, the CI at any given time(s) can be displayed with `summary`:

```
summary(fit, times=c(8, 10, 16))
```

- By default, `plot(fit)` will plot the pointwise CIs if there is only one group; otherwise you have to request them with

```
plot(fit, conf.int=TRUE)
```

## Shaded CIs

- It should be noted that the `survival` package depicts CIs using dotted lines; as you can probably tell, I think shaded bands are much clearer and better-looking
- Unfortunately, there is no option in the `survival` package to get them
- If you want them, you'll either have to use a different R package, or the `ciband` function I've provided, which works well with the `Plot` function we introduced last time:

```
Plot(fit)  
ciband(fit)
```

## CIs for the median survival time

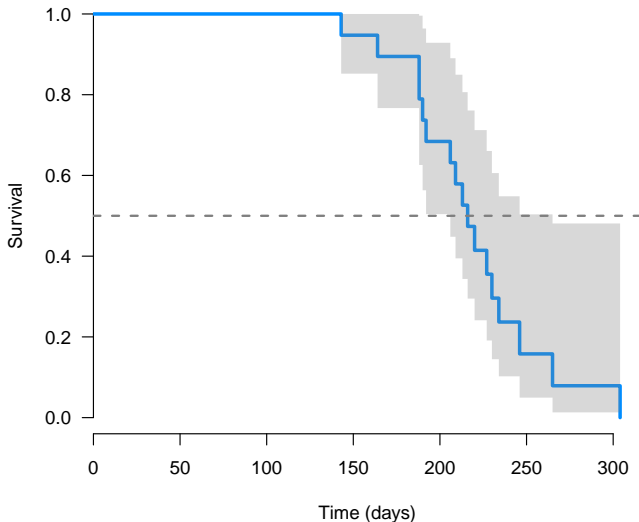
- Obtaining confidence intervals for  $S(t)$  also allows us to obtain confidence intervals for the median survival time (or any other quantile of interest)
- Because of the duality between confidence intervals and hypothesis tests, we can think of whether the CI for  $S(t)$  contains 0.5 as a test at  $\alpha = 0.05$  of  $H_0 : S(t) = 0.5$
- A reasonable confidence interval for the median is therefore the set of all such times for which the above hypothesis is not rejected

## Pike (1966) data

- Because our GVHD data does not permit the estimation of median survival, let's switch to some data from a carcinogenesis experiment reported by Pike (1966)
- The experiment consisted of exposing female rats to the carcinogen DMBA and observing the progression of disease until the rats died of vaginal cancer
- The original study was divided into two groups on the basis of a pretreatment regimen, but we'll restrict our attention to group 1 for today



## Pike (1966): CI for median survival time



## CIs for median survival with `survfit`

This confidence interval can be obtained from the output of `survfit` as follows:

```
> fit$time[ $\min(\text{which}(\text{fit}\$lower < 0.5))$ ]  
[1] 206  
> fit$time[ $\max(\text{which}(\text{fit}\$upper > 0.5))+1$ ]  
[1] 265
```

Although keep in mind that you also get this interval by just printing the `survfit` object:

```
> fit  
      n  events  median 0.95LCL 0.95UCL  
19     17     216     206     265
```

# The Nelson-Aalen estimator

- Finally, it's worth mentioning, given the central role of hazard functions in survival, that many authors consider the estimation of the cumulative hazard to be the more “natural” thing to estimate than the survival function
- Following the same lines as we did earlier, the nonparametric MLE for the cumulative hazard is

$$\begin{aligned}\hat{\Lambda}(t) &= \sum_{t_j \leq t} \hat{\lambda}_j \\ &= \sum_{t_j \leq t} d_j/n_j;\end{aligned}$$

this is known as the *Nelson-Aalen estimator*

# Standard errors for the Nelson-Aalen estimator

Obtaining standard errors for the Nelson-Aalen estimator is straightforward, since it is simply a sum of binomial probability estimates:

$$\begin{aligned}\mathbb{V}\hat{\Lambda}(t) &= \sum_{t_j \leq t} \mathbb{V}\hat{\lambda}_j \\ \implies \hat{\mathbb{V}}\hat{\Lambda}(t) &= \sum_{t_j \leq t} \frac{d_j(n_j - d_j)}{n_j^3}\end{aligned}$$

# Nelson-Aalen estimator

