Nonparametric methods: hazard, cumulative hazard, and the Nelson-Aalen estimator of $S(t)$

Shannon Pileggi

STAT 417

OUTLINE

Estimating the hazard function

- Recall that if we want to assess the instantaneous risk of failure (experiencing the event) at time t given survival (not experiencing the event) to time t, we can use the **hazard function** $h(t)$.
- We can estimate hazard rates using a two approaches:
- **Nelson-Aalen Type** $\tilde{h}(t)$: The estimated hazard rate at time $t_{(i)}$, $i = 0, \ldots, m-1$, denoted $\tilde{h}(t_{(i)})$ is given by:

and the estimated rate at time $t_{(m)}$ is $\tilde{h}(t_{(m)}) = d_m/n_m$.

Estimating the hazard function

- **Kaplan-Meier Type** $\hat{h}(t)$: The estimated hazard rate at time $t_{(i)}$, $i = 0, \ldots, m-1$, denoted $\hat{h}(t_{(i)})$ is given by:

and the estimated rate at time $t_{(m)}$ is $\hat{h}(t_{(m)}) = \hat{h}(t_{(m-1)})$.

Discussion

Which estimator do you think more closely aligns with the definition of hazard?

1. Nelson-Aalen Type
2. Kaplan-Meier Type

Why?

Interpretations of hazard estimators

- **Interpretation of** $\tilde{h}(t_{(i)})$:
- **Interpretation of** $\hat{h}(t_{(i)})$:

Estimating hazard

i	Interval	n_i	d_i	$n_i - d_i$	$\widehat{S}(t)$	$\tilde{h}(t)$	$\hat{h}(t)$
0	$[0, 1.41)$	7	0	7	1		
1	$[1.41, 3.56)$	7	1	6	.857		
2	$[3.56, 4.18)$	4	1	3	.643		
3	$[4.18, 13.18)$	3	1	2	.429		
4	$[13.18, 13.18]$	1	1	0	0		

Interpreting hazard

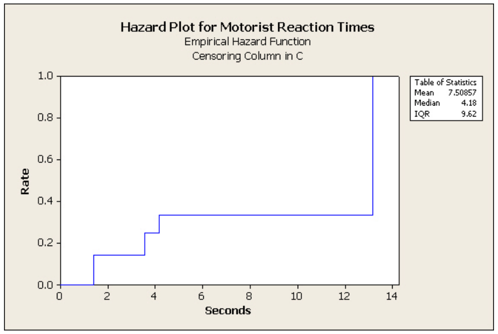
Interpret $\tilde{h}(t_{(1)})$ and $\hat{h}(t_{(1)})$.

Graphing the Estimated Hazard Functions

By examining the graph of the estimated hazard function, we can investigate how the risk of event occurrence changes over time.

- Similar to the Kaplan-Meier curve, the graphs of $\tilde{h}(t)$ and $\hat{h}(t)$ are step functions with steps occurring at each complete event time.
- The estimated hazard curves are 0 outside the smallest and largest complete event times.
- Minitab only plots the Nelson-Aalen type estimator $\tilde{h}(t)$. Other software must be used to produce a plot of $\hat{h}(t)$ (will use the R software).
- If the largest observed time is complete, then $\tilde{h}(t_{(m)}) = 1$.

Minitab plot of $\tilde{h}(t)$

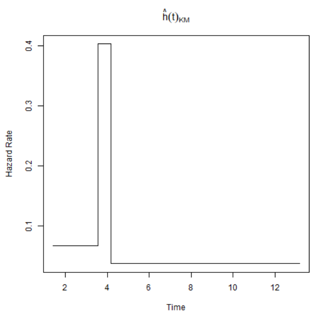


R plot of $\hat{h}(t)$

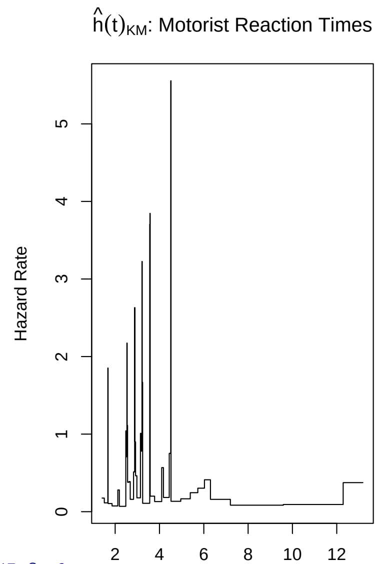
R Code

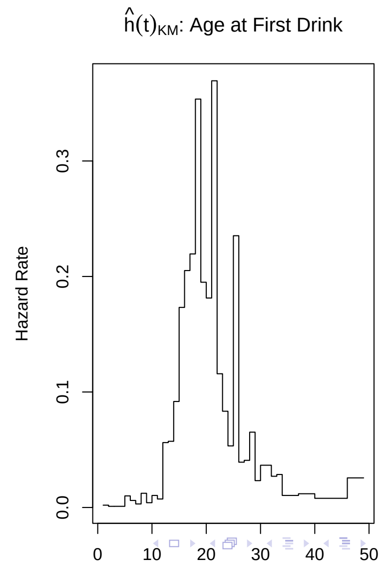
```
time <- c(1.41, 1.41, 2.76, 3.56, 4.18, 4.71, 13.18)
censor <- c(1, 0, 0, 1, 1, 0, 1)
KM_obj <- survfit(Surv(time, censor) ~ 1)
plot_haz(KM_obj)  # user written function
```

R Code



Two example hazards





Observations from two example hazards

Cumulative hazard function

- The coarse nature of the estimated hazard curve $\hat{h}(t)$ can make it difficult to describe or summarize how the conditional risk of event occurrence changes over time.
- Recall during our discussion of parametric models that an alternative method to assess and describe how the parametric hazard function $h(t)$ changes over time is to investigate the accumulation of the hazard rates over time, i.e. examine the *cumulative hazard function*, $H(t)$.
- The cumulative hazard function $H(t)$ is an accumulation of the population hazard $h(t)$ between time 0 and t:
- Then an estimator of $H(t)$ should also accumulate (or sum up) the estimated hazard rates computed between time 0 and time t.

Nelson-Aalen type estimator of the cumulative hazard function

An estimator of $H(t)$ can be derived from the Nelson-Aalen type estimates $\tilde{h}(t_{(i)})$ for $i = 1, \ldots, m$:

- Recall that at time $t_{(i)}$, $i = 0, \ldots, m-1$:
$$\tilde{h}(t_{(i)}) = d_i/n_i$$
with $\tilde{h}(t_{(m)}) = d_m/n_m$.
- So the *Nelson-Aalen* estimator of $H(t)$, denoted $\tilde{H}(t)$, is simply the sum of these total estimated hazard quantities up to a particular time t, given by:

Alternative Estimator of $H(t)$

- Recall the relationship between the (population) cumulative hazard function $H(t)$ and the (population) survival function $S(t)$:
- This result provides an alternative estimator for the cumulative hazard function, denoted $\widehat{H}(t)$:
- Since this estimator of $H(t)$ is based $\widehat{S}(t)$, it is sometimes referred to as the Kaplan-Meier estimator of $H(t)$.
- $\widehat{H}(t)$ has worse small sample size performance than $\tilde{H}(t)$ so it is not used in practice as often.

Estimating cumulative hazard

i	Time Interval	$\widehat{S}(t)$	$\tilde{h}(t)$	$\hat{h}(t)$	$\tilde{H}(t)$	$\widehat{H}(t)$
0	$[0, 1.41)$	1	0	0		
1	$[1.41, 3.56)$	.857	.143	.066		
2	$[3.56, 4.18)$	.643	.250	.403		
3	$[4.18, 13.18)$	.429	.333	.037		
4	$[13.18, 13.18]$	0	1	.037		

Details of the Graph of the Nelson-Aalen Estimator $\tilde{H}(t)$

The graph of $\tilde{H}(t)$ is a step function with steps occurring at the complete times, and can be used to summarize changes in the estimated hazard rate over time.

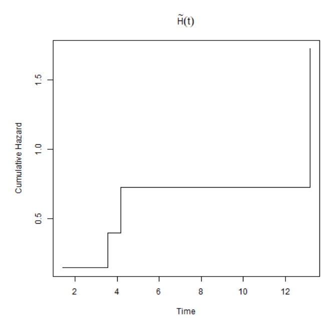
- If the last observed event time is censored, then $\tilde{H}(t)$ will peak (reach its highest point) at the largest complete event time, $t_{(m)}$, and then extend to the largest censored event time, $t_{\max}$, i.e. it will be constant over the interval $[t_{(m)}, t_{\max}+)$.
- If the last observed event time is complete, then $\tilde{H}(t)$ will simply reach its highest value at the complete time, $t_{(m)}$.
- Graphs of $\tilde{H}(t)$ are not available in Minitab, and should be constructed using R.

R Plot of $\tilde{H}(t)$

R Code

```
time <- c(1.41, 1.41, 2.76, 3.56, 4.18, 4.71, 13.18)
censor <- c(1, 0, 0, 1, 1, 0, 1)
KM_obj <- survfit(Surv(time, censor) ~ 1)
plot_chaz(KM_obj)  # user written function
```

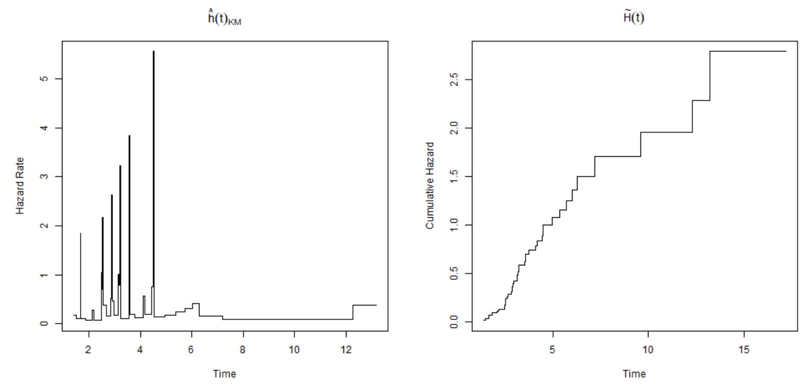
R Code



Relationship between $\hat{h}(t)$ and $\tilde{H}(t)$

- If the *rate* of increase in $\tilde{H}(t)$ is *increasing* (over an interval of time), then:
- If the *rate* of increase in $\tilde{H}(t)$ is *decreasing*, then:
- If the *rate* of increase in $\tilde{H}(t)$ is *constant* (and greater than zero), then:
- If the *rate* of increase in $\tilde{H}(t)$ is 0, then:

$\hat{h}(t)$ and $\tilde{H}(t)$ (all motorist reaction times)



$\hat{h}(t)$ and $\tilde{H}(t)$ (all motorist reaction times)

Alternative estimator of $S(t)$: Nelson-Aalen estimator

- Once again recall the relationship between $S(t)$ and $H(t)$:
  $$H(t) = -\ln[S(t)]$$
- We can solve for the survival function $S(t)$:
- Then using the Nelson-Aalen estimator of the cumulative hazard function, $\tilde{H}(t)$, we can derive the **Nelson-Aalen estimator of** $S(t)$, denoted $\tilde{S}(t)$, given by:

Nelson-Aalen estimator of $S(t)$

i	Time Interval	$\tilde{h}(t)$	$\tilde{H}(t)$	$\tilde{S}(t)$	$\widehat{S}(t)$
0	$[0, 1.41)$	0	0		1
1	$[1.41, 3.56)$	.143	.143		.857
2	$[3.56, 4.18)$	.250	.393		.643
3	$[4.18, 13.18)$	.333	.726		.429
4	$[13.18, 13.18]$	1	1.726		0

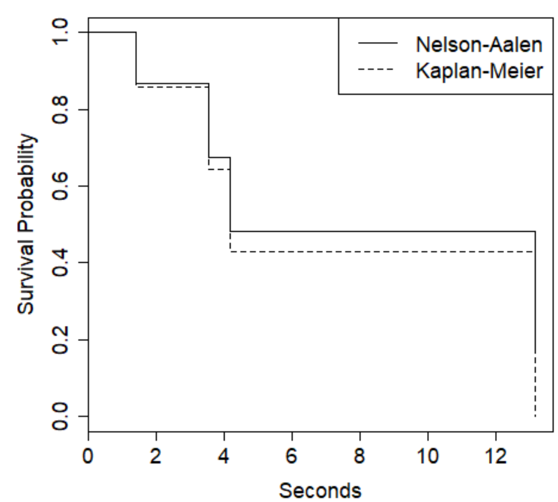
Calculation of $\tilde{S}(t)$ in R

R Code

```
time <- c(1.41, 1.41, 2.76, 3.56, 4.18, 4.71, 13.18)
censor <- c(1, 0, 0, 1, 1, 0, 1)

KM_obj_na <- survfit(Surv(time, censor) ~ 1,
                     type = "fh",
                     conf.type = "none")

KM_obj_km <- survfit(Surv(time, censor) ~ 1,
                     conf.type = "none")

plot(KM_obj_na, xlab="Seconds", ylab="Survival Probability")
lines(KM_obj_km, lty=2)
legend("topright",c("Nelson-Aalen","Kaplan-Meier"),lty=1:2)
```

R Code

R Plot of $\tilde{S}(t)$ and $\widehat{S}(t)$



Summary of nonparametric methods

- Kaplan-Meier type estimators:
  1. Survival function: $\widehat{S}(t)$
  2. Hazard function: $\hat{h}(t)$ (R required)
  3. Cumulative hazard function: $\hat{H}(t) = -\ln[\widehat{S}(t)]$ (R required)
- Nelson-Aalen type estimators:
  1. Survival function: $\tilde{S}(t) = \exp[-\tilde{H}(t)]$ (R required)
  2. Hazard function: $\tilde{h}(t)$
  3. Cumulative hazard function: $\tilde{H}(t)$ (R required)
- Descriptive measures (using the Kaplan-Meier estimator $\widehat{S}(t)$):
  1. Estimated mean survival time: $\hat{\mu}$
  2. Estimated percentiles of survival time: $\hat{t}_p$