

Aalen's Additive Regression

Eva Burns

12/12/2021

This will show the output of Aalen's regression over different types of data and censoring. Unfortunately, we did not have time to build a model for Aalen's, so there will be no model evaluations.

Import Libraries and External Files

```
library(survival)
library(ggplot2)
library(dplyr)
library(ggfortify)
library(gridExtra)

source("../SimulationFunctions.R")
source("../convert_list.R")
```

Simulate Data

```
# hollings
data_hollings <- sim_hollings(1000)
nat_hollings <- data_hollings$nat
fvt_hollings <- data_hollings$fvt

# sin with linear relationship to hazard
data_sin <- sim_sin(1000)
nat_sin <- data_sin$nat
fvt_sin <- data_sin$fvt

# sin with exponential relationship to hazard
data_exp_sin <- sim_exp_sin(1000)
nat_exp_sin <- data_exp_sin$nat
fvt_exp_sin <- data_exp_sin$fvt
```

Covert Data to Readable Format

The data currently has one row per organism and has the trait values until they die, which after is just NAs. For Aalen's, each measurement should be one one row.

```
nat_hollings <- convert_list(nat_hollings)
nat_hollings$time0 = nat_hollings$time - 1 # add extra time col for interval censoring
fvt_hollings <- convert_list(fvt_hollings)
fvt_hollings$time0 = fvt_hollings$time - 1

nat_sin <- convert_list(nat_sin)
```

```

nat_sin$time0 = nat_sin$time - 1
fvt_sin <- convert_list(fvt_sin)
fvt_sin$time0 = fvt_sin$time - 1

nat_exp_sin <- convert_list(nat_exp_sin)
nat_exp_sin$time0 = nat_exp_sin$time - 1
fvt_exp_sin <- convert_list(fvt_exp_sin)
fvt_exp_sin$time0 = fvt_exp_sin$time - 1

```

Here's a comparison of the data:

```

# Before converting
head(data_hollings$nat, 1)

```

```

##      [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]      [,8]
## [1,]    0 7.083249 13.00552 18.03063 22.34809 26.09754 29.38417 32.28867
##      [,9]      [,10]     [,11]     [,12]     [,13]     [,14]     [,15] [,16] [,17]
## [1,] 34.87404 37.19013 39.27693 41.16689 42.88659 44.45806 45.89966    NA    NA
##      [,18] [,19] [,20] [,21] [,22] [,23] [,24] [,25] [,26] [,27] [,28] [,29]
## [1,]    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA
##      [,30] [,31] [,32] [,33] [,34] [,35] [,36] [,37] [,38] [,39] [,40] [,41]
## [1,]    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA
##      [,42] [,43] [,44] [,45] [,46] [,47] [,48] [,49] [,50] [,51] [,52] [,53]
## [1,]    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA
##      [,54] [,55] [,56] [,57] [,58] [,59] [,60] [,61] [,62] [,63] [,64] [,65]
## [1,]    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA
##      [,66] [,67] [,68] [,69] [,70] [,71] [,72] [,73] [,74] [,75] [,76] [,77]
## [1,]    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA
##      [,78] [,79] [,80] [,81] [,82] [,83] [,84] [,85] [,86] [,87] [,88] [,89]
## [1,]    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA
##      [,90] [,91] [,92] [,93] [,94] [,95] [,96] [,97] [,98] [,99] [,100] [,101]
## [1,]    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA
##      [,102] [,103] [,104] [,105] [,106] [,107] [,108] [,109] [,110] [,111]
## [1,]    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA
##      [,112] [,113] [,114] [,115] [,116] [,117] [,118] [,119] [,120] [,121]
## [1,]    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA
##      [,122] [,123] [,124] [,125] [,126] [,127] [,128] [,129] [,130]
## [1,]    NA    NA    NA    NA    NA    NA    NA    NA    NA

```

```

# After converting
head(nat_hollings, 5)

```

```

##   time status   trait ids time0
## 1     1      0 7.083249     1     0
## 2     2      0 13.005521     1     1
## 3     3      0 18.030631     1     2
## 4     4      0 22.348089     1     3
## 5     5      0 26.097544     1     4

```

Right Censored Analysis

Using right censored data on Aalen's regression. We will only look at the output of Hollings type II in depth.

Hollings:

Does not affect survival

```
aa_hollings_nat_right <- aareg(Surv(time, status) ~ trait, data = nat_hollings)
aa_hollings_nat_right
```

Call:

```
## aareg(formula = Surv(time, status) ~ trait, data = nat_hollings)
```

##

```
## n=20420 (56 observations deleted due to missingness)
```

```
## 94 out of 95 unique event times used
```

##

```
##          slope      coef se(coef)      z      p
```

```
## Intercept 0.18800 4.82e-04 2.00e-05 24.1 4.28e-128
```

```
## trait     -0.00275 -9.98e-06 3.99e-07 -25.0 6.51e-138
```

##

```
## Chisq=624.87 on 1 df, p=<2e-16; test weights=aalen
```

Affects survival

```
aa_hollings_fvt_right <- aareg(Surv(time, status) ~ trait, data = fvt_hollings)
```

```
aa_hollings_fvt_right
```

Call:

```
## aareg(formula = Surv(time, status) ~ trait, data = fvt_hollings)
```

##

```
## n=11908 (48 observations deleted due to missingness)
```

```
## 52 out of 53 unique event times used
```

##

```
##          slope      coef se(coef)      z      p
```

```
## Intercept 0.3750 7.20e-04 2.94e-05 24.5 4.49e-132
```

```
## trait     -0.0068 -1.82e-05 7.13e-07 -25.6 2.89e-144
```

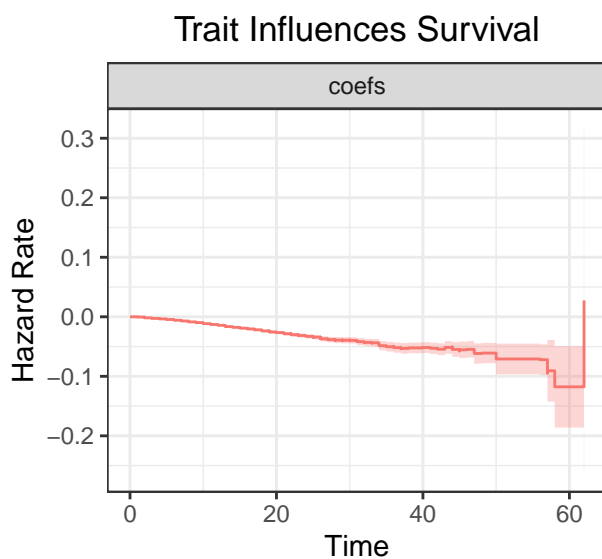
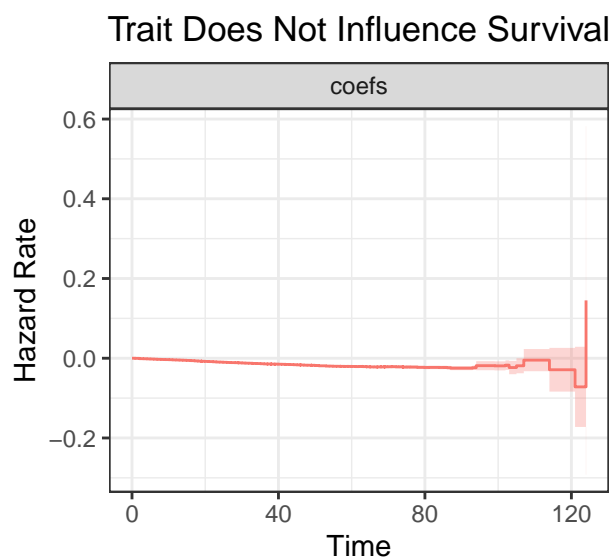
##

```
## Chisq=654.09 on 1 df, p=<2e-16; test weights=aalen
```

```
plot1 <- autoplot(aa_hollings_nat_right[2], ylab="Hazard Rate", xlab="Time", main="Trait Does Not Infl
```

```
plot2 <- autoplot(aa_hollings_fvt_right[2], ylab="Hazard Rate", xlab="Time", main="Trait Influences Su
```

```
grid.arrange(plot1, plot2, ncol=2)
```



Sin with linear relationship to hazard:

```
# Does not affect survival
aa_sin_nat_right <- aareg(Surv(time, status) ~ trait, data = nat_sin)

# Affects survival
aa_sin_fvt_right <- aareg(Surv(time, status) ~ trait, data = fvt_sin)
```

Sin with exponential relationship to hazard:

```
# Does not affect survival
aa_exp_sin_nat_right <- aareg(Surv(time, status) ~ trait, data = nat_exp_sin)

# Affects survival
aa_exp_sin_fvt_right <- aareg(Surv(time, status) ~ trait, data = fvt_exp_sin)
```

The table below summarizes the output of right censored Aalen's on all of the types of simulated data.

		Slope	Coef	P-value
Hollings	Doesn't influence survival	-0.00275	10^{-5}	0
	Influences survival	-0.0068	-2×10^{-5}	0
Sin (linear)	Doesn't influence survival	0.00134	10^{-5}	0.06
	Influences survival	6.3×10^{-4}	0	0.618
Sin (exponential)	Doesn't influence survival	-9.7×10^{-4}	10^{-5}	0.138
	Influences survival	-5.8×10^{-4}	10^{-5}	0.418

The Hollings type II data was the only type that had a consistently low enough p-value to make a conclusion about the slope and coefficient. Running this code multiple times on the sin data will give different results every time. The slope and coefficient of the Hollings data when the trait influences survival typically was lower than if the trait did not influence survival.

Interval Censored Analysis

Using interval censored data on Aalen's regression. We will only look at the output of Hollings type II in depth.

Hollings:

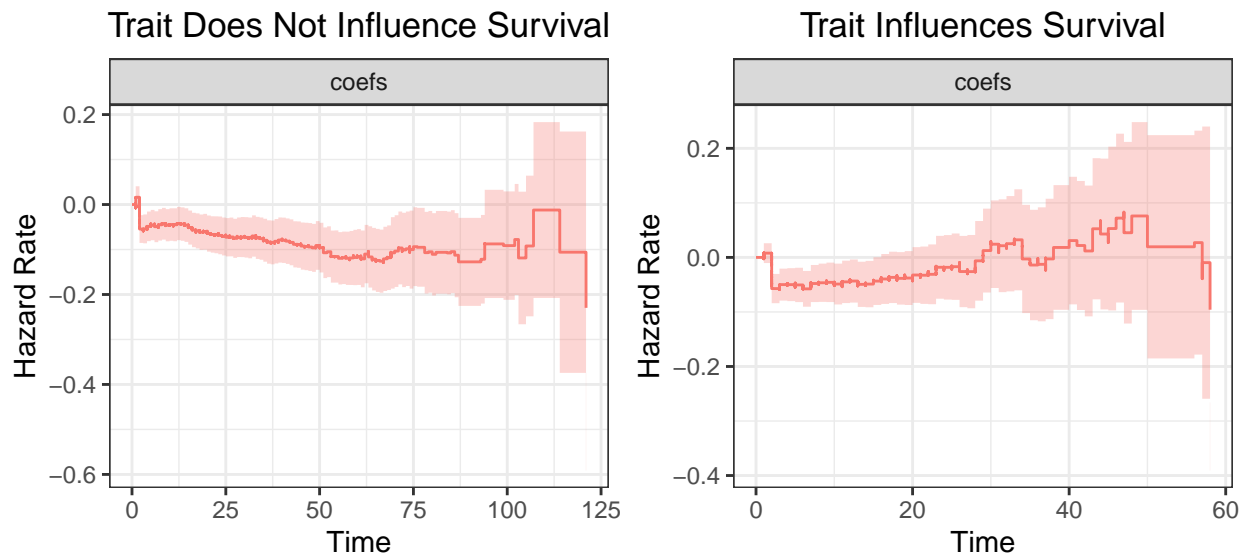
```
# Does not affect survival
aa_hollings_nat_interval <- aareg(Surv(time0, time, status) ~ trait, data = nat_hollings)
aa_hollings_nat_interval
```

```
## Call:
## aareg(formula = Surv(time0, time, status) ~ trait, data = nat_hollings)
##
##      n=20420 (56 observations deleted due to missingness)
##      93 out of 95 unique event times used
##
##              slope      coef se(coef)      z      p
## Intercept  1.8900  0.006690 7.98e-04  8.38 5.15e-17
## trait      -0.0086 -0.000185 3.01e-05 -6.14 8.06e-10
##
## Chisq=37.75 on 1 df, p=8.06e-10; test weights=aalen
```

```
# Affects survival
aa_hollings_fvt_interval <- aareg(Surv(time0, time, status) ~ trait, data = fvt_hollings)
aa_hollings_fvt_interval
```

```
## Call:
## aareg(formula = Surv(time0, time, status) ~ trait, data = fvt_hollings)
##
##      n=11908 (48 observations deleted due to missingness)
##      51 out of 53 unique event times used
##
##              slope      coef se(coef)      z      p
## Intercept  2.1800  0.005300 0.000774  6.85 7.34e-12
## trait      -0.0129 -0.000106 0.000034 -3.11 1.88e-03
##
## Chisq=9.66 on 1 df, p=0.00188; test weights=aalen
```

```
plot1 <- autoplot(aa_hollings_nat_interval[2], ylab="Hazard Rate", xlab="Time", main="Trait Does Not Influence Survival")
plot2 <- autoplot(aa_hollings_fvt_interval[2], ylab="Hazard Rate", xlab="Time", main="Trait Influences Survival")
grid.arrange(plot1, plot2, ncol=2)
```



Sin with linear relationship to hazard:

```
# Does not affect survival
aa_sin_nat_interval <- aareg(Surv(time0, time, status) ~ trait, data = nat_sin)

# Affects survival
aa_sin_fvt_interval <- aareg(Surv(time0, time, status) ~ trait, data = fvt_sin)
```

Sin with exponential relationship to hazard:

```
# Does not affect survival
aa_exp_sin_nat_interval <- aareg(Surv(time0, time, status) ~ trait, data = nat_exp_sin)

# Affects survival
aa_exp_sin_fvt_interval <- aareg(Surv(time0, time, status) ~ trait, data = fvt_exp_sin)
```

The table below summarizes the output of interval censored Aalen's on all of the types of simulated data.

		Slope	Coef	P-value
Hollings	Doesn't influence survival	-0.0086	-1.8×10^{-4}	0
	Influences survival	-0.01287	-1.1×10^{-4}	0.002
Sin (linear)	Doesn't influence survival	0.01969	2.2×10^{-4}	0.225
	Influences survival	-0.00202	-9×10^{-5}	0.608
Sin (exponential)	Doesn't influence survival	-0.0164	-2×10^{-4}	0.263
	Influences survival	-0.0117	-1.3×10^{-4}	0.458

The Hollings type II data was the only type that had a consistently low enough p-value to make a conclusion about the slope and coefficient. Running this code multiple times on the sin data will give different results every time. The slope of the Hollings data when the trait influences survival typically was lower than if the trait did not influence survival.

In conclusion, only the Hollings type II data worked well with Aalen's regression. Typically when the trait influences survival, the slope is lower than if it did not influence survival.