Accelerated Failure Time Models

Matthew Austin

Institute for Advanced Analytics
MSA Class of 2019

AFT model:
Overview

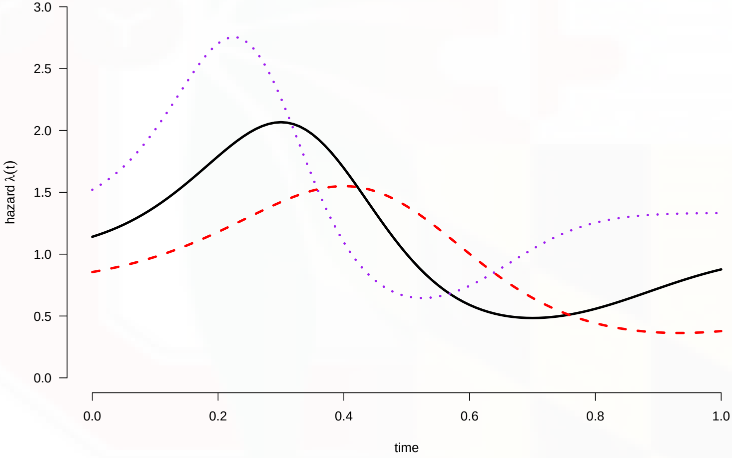
Regression model for survival data

- We are often interested in modeling the relationship between predictors and a response, but survival data is challenging to do with standard modeling techniques
  - Censored data
  - Time-to-event data
  - Distribution of time
- But, we can make a linear model as long as we specify the underlying distribution of time

"Underlying distribution of time?"

- We're familiar with many examples of how different things affect or "rescale" time
  - "Dog years"
  - "A New York minute"
- The idea is that we have some preconceived notion of how time behaves in normal/baseline circumstances—a **distribution**—but certain factors change the particular amount or length of time
- Thus, the basis of our model is that the predictors affect the amount by which time is rescaled in some way

A rescaled hazard distribution



AFT model:
Framework

Accelerated failure time model

- An **accelerated failure time model** (AFT) proposes a linear relationship between predictors and log T:

$$\log T_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \sigma \varepsilon_i$$

- Then on T directly, the model is:

$$T_i = T_0 e^{\mathbf{x}_i^\top \boldsymbol{\beta}},$$

  where $T_0 = e^{\sigma \varepsilon_i}$

- So in an AFT model, an individual's predictor values $\mathbf{x}_i$ rescale the normal/baseline time T_0 by some amount $e^{\mathbf{x}_i^\top \boldsymbol{\beta}}$

A quick note on notation

- Whenever we have a linear model of the form $\mathbf{x}_i^\top \boldsymbol{\beta} =$ (something), the quantity $\mathbf{x}_i^\top \boldsymbol{\beta}$ is called the **linear predictor**
- I will sometimes represent linear predictors by η
  - In linear regression, $\eta_i = y_i = \mathbf{x}_i^\top \boldsymbol{\beta}$
  - In logistic regression, $\eta_i = \log(\text{odds}) = \mathbf{x}_i^\top \boldsymbol{\beta}$, so $p_i = \frac{e^{\eta_i}}{1+e^{\eta_i}}$
  - In AFT models, $\eta_i = \log T_i = \mathbf{x}_i^\top \boldsymbol{\beta}$, so $T_i = T_0 e^{\eta_i}$

Effect on survival time

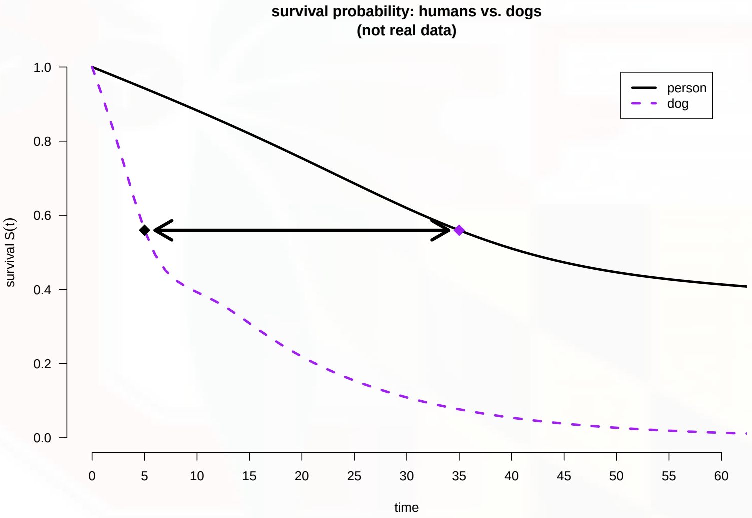
$$T_i = T_0 e^{\mathbf{x}_i^\top \boldsymbol{\beta}} = T_0 e^{\eta_i}$$

- In an AFT, notice that we are multiplying the baseline event time T_0 by $e^{\eta_i} = e^{\mathbf{x}_i^\top \boldsymbol{\beta}}$
- If $e^{\eta_i} = 1$, then the individual doesn't experience time any differently than normal
- If $e^{\eta_i} = 1/2$, then the individual's failure time is *sooner*—they "age" twice as fast as normal
- If $e^{\eta_i} = 2$, then the individual's failure time is *later*—they "age" at half the normal speed

Acceleration factor

- In AFTs, rescaling time also means that survival probabilities and hazards are rescaled
    - $S_i(t) = S_0(te^{-\eta_i})$
    - $\lambda_i(t) = \lambda_0(te^{-\eta_i})e^{-\eta_i}$
- $e^{-\eta_i}$ is called the **acceleration/aging factor** (notice the negative sign!)
- Example: if the acceleration factor $e^{-\eta_i} = 2$, then individual i's survival probability at `time = 9` is $S_i(9) = S_0(2\times 9) = S_0(18)$
- Dogs [allegedly] age $7x$ faster than humans, so $e^{\eta} = 1/7$, the acceleration factor is 7, and $S_{\texttt{dog}}(t) = S_{\texttt{human}}(7t)$

Survival: “dog years”



Interpreting estimates

- $e^{c_j\beta_j}$ is the predicted change in T_i for $x_j + c_j$ compared to x_j
  - $\beta_j = 0 \iff e^{\beta_j} = 1$ means no change in predicted failure time as x_j increases
  - $\beta_j > 0 \iff e^{\beta_j} > 1$ means predicted failure time *increases* (survives longer) as x_j increases
  - $\beta_j < 0 \iff e^{\beta_j} < 1$ means predicted failure time *decreases* (dies sooner) as x_j increases

AFT models: R syntax

```
fit <- survreg(Surv(week, arrest == 1) ~ fin + age +
                  race + wexp + mar + paro + prio,
               data = recid, dist = "weibull")
```

	Value	Std. Error	z	p
(Intercept)	3.99	0.42	9.52	0.00
fin	0.27	0.14	1.97	0.05
age	0.04	0.02	2.54	0.01
race	-0.22	0.22	-1.02	0.31
wexp	0.11	0.15	0.70	0.48
mar	0.31	0.27	1.14	0.25
paro	0.06	0.14	0.42	0.67
prio	-0.07	0.02	-3.14	0.00
Log(scale)	-0.34	0.09	-3.81	0.00

Interpretation

	Value	exp_Value
(Intercept)	3.99	54.06
fin	0.27	1.31
age	0.04	1.04
race	-0.22	0.80
wexp	0.11	1.11
mar	0.31	1.37
paro	0.06	1.06
prio	-0.07	0.94
Log(scale)	-0.34	0.71

- For men who received financial aid, the predicted recidivism time was $\beta_{\texttt{fin}} = e^{0.27} = 1.31$ times longer than those who didn't
- The predicted time to recidivism is $\beta_{\texttt{prio}} = e^{-0.07} = 0.94$ times shorter for every additional prior conviction
- (All else equal, etc.)

AFT models: SAS syntax

```
proc lifereg data=survival.recid;
  model week*arrest(0) = fin age race wexp mar prio paro;
run;
```

Error distributions:
Model assumptions

Parametric models

- AFT models are **parametric**—we assume failure time has a particular structure and distribution
- cf. Kaplan-Meier estimation, a *nonparametric* method which makes no assumptions about T
- Parametric methods allow for more detailed/precise estimation than nonparametric methods if the distribution is specified correctly
  - Estimating median survival times (or other quantiles)
  - Estimating expected/mean event times
  - Estimating survival and hazard curves

"…If the distribution is specified correctly"

$$\log T_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \sigma \varepsilon_i$$

- ε_i are errors that follow some particular distribution (probably not a normal distribution!)
- σ is a **scale** parameter that determines the spread of ε_i
- Assumptions:
  - Linear relationship between x and $\log T$
  - **The error distribution is specified correctly**
  - σ and $\mathbf{x}$ are independent

Scale vs. rate

- A scale parameter σ describes the spread of a distribution (like variance or standard deviation)
- A related idea is the **rate** parameter $\lambda = 1/\sigma$
- If σ is small (thus λ is large), then the event times aren't spread out, so events are happening close to one another; i.e., at a higher rate
- The rate is probably the more natural thing to use in survival analysis, as the hazard is really just telling us the risk of failure in some interval—the rate at which events are happening (conditional on it not having already happened)
- I prefer to use the rate whenever possible for this reason, but as we will see, both are commonly (and inconsistently) used

Error distributions:
Common distributions

Distribution of...?

- Note that when I say "distribution" here, that means the distribution of T and e^{ε}, not $\log T$ and ε, and this is what you are telling R and SAS
- Distributions can be checked graphically or through testing
- We will go over some commonly used distributions for survival data, but there's no guarantee that your data will adequately match just one of the distributions discussed here, or even any of them at all

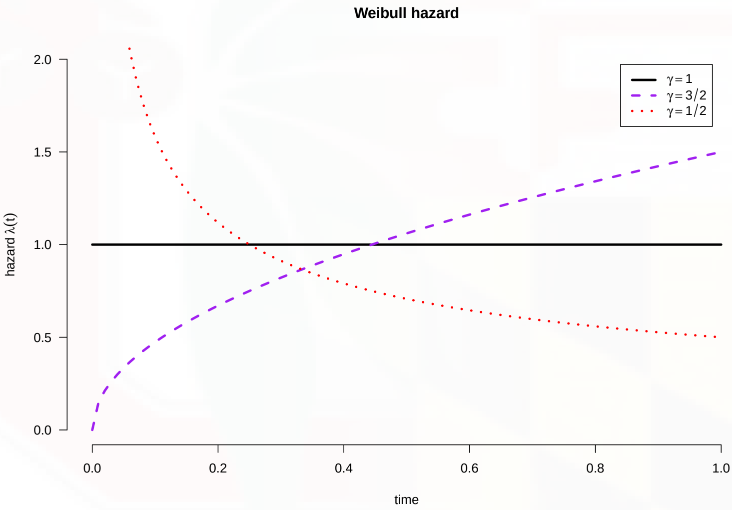
The exponential distribution

- The simplest distribution is the **exponential distribution**, which has a constant hazard that doesn't depend on time:
    - $\lambda(t) = \lambda$
    - $S(t) = e^{-\lambda t}$
    - $\Lambda(t) = \lambda t$
- A constant hazard is commonly used when failures are completely random (light bulbs, electronics, etc.)

The Weibull distribution

- Probably the most common distribution is the **Weibull distribution**, which has an additional *shape* parameter γ
  - $\lambda(t) = \lambda\gamma(\lambda t)^{\gamma-1}$
  - $S(t) = \exp\{-(\lambda t)^{\gamma}\}$
  - $\Lambda(t) = (\lambda t)^{\gamma}$
- γ determines whether the hazard increases/decreases with time:
  - When $\gamma > 1$, then the hazard is increasing with time (ex: aging parts, "wear-out", etc.)
  - When $\gamma < 1$, then the hazard is decreasing with time (ex: post-surgery complications, etc.)

Weibull hazards



Relationship between exponential and Weibull

- On the previous slide, notice that when $\gamma = 1$, the hazard is constant...
- ...so the Weibull distribution with $\gamma = 1$ is the exponential distribution
- `R` and `SAS` automatically test this when `dist = "weibull"` or `dist=exponential`, respectively, are specified
    - In `R`, the `Log(scale)` line (wait a minute... what?) is an estimate of $\log(1/\gamma)$ and associated p-value is for testing if it equals 0 (which is equivalent to testing $H_0: \gamma = 1$)
    - In `SAS`, the "Lagrange multiplier test for scale" (...seriously?) is testing $H_0: \gamma = 1$

A note on the parameterization

- With the scale vs. rate or shape vs. scale thing, there are a couple different ways to write the Weibull distribution, and they're all fairly common, as stated previously
  - `?survreg`: "There are multiple ways to parameterize a Weibull distribution. The `survreg` function embeds it in a general location-scale family, which is a different parameterization than the `rweibull` function, and often leads to confusion."
  - `proc lifereg` documentation: "The Weibull with *Scale=1* is an exponential distribution."
- (See course page if you care about why)

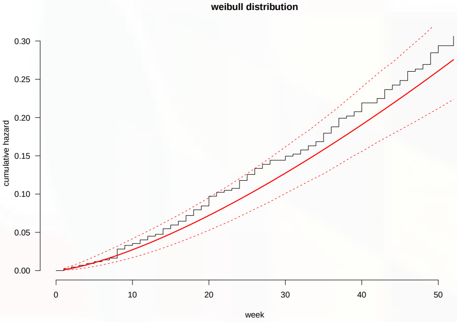
Matching up the parameterization

- `proc lifereg` "Weibull Shape" = γ
- `survreg` "scale" = `proc lifereg` "scale" = $1/\gamma$
- `survreg` "intercept" = `proc lifereg` "intercept" = $-\log\lambda$
- (Again, see course page if you care about why)

Checking distributions: R syntax

Unfortunately, plotting `survreg` objects isn't straightforward, so we'll use the `flexsurv` package to check distributions using plots

```
fit_wb <- flexsurvreg(...)
plot(fit_wb, type = "cumhaz")
```



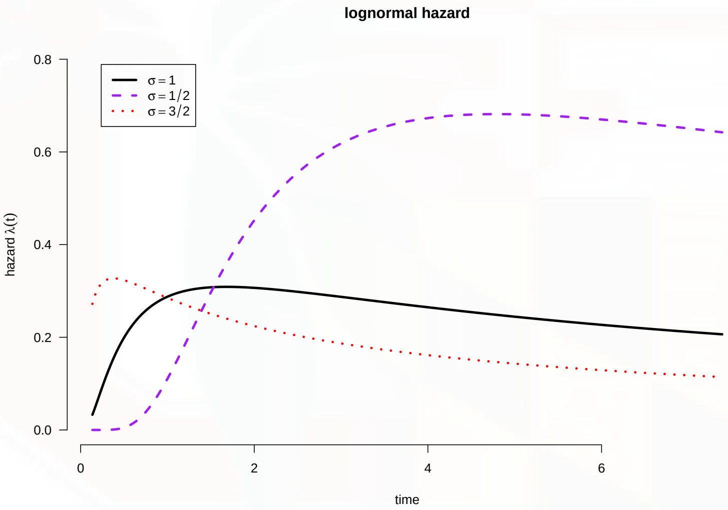
Checking distributions: SAS syntax

```
proc lifereg data=survival.recid;
  model week*arrest(0) = fin age race wexp mar prio paro;
  probplot;
run;
```

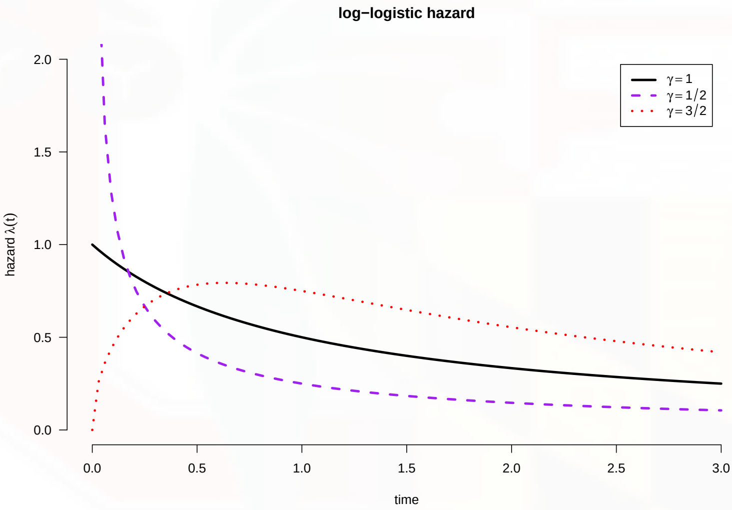
Other distributions

- The **lognormal distribution**: if T has a lognormal distribution, then ε follows a normal distribution
  - If you pretend there's no censoring, a lognormal AFT and linear regression with $y = \log T$ are equivalent
- The **log-logistic distribution**: can allow hazard to increase then decrease if $\gamma > 1$
  - The log-logistic AFT is just an ordinal logistic regression model! (again, see course page if you care about why)
- Generalized Gamma: includes lognormal and Weibull as special cases
- Generalized F: includes generalized gamma and log-logistic as special cases

Lognormal hazard



Log-logistic hazard



Error distributions:
Predicting survival/event times

Making predictions

- In AFT models, we assume a distribution for T, meaning that we expect event times to behave in a certain way
- That might be an unreasonable assumption to make, but if it isn't, then the distribution allows us to predict quantiles, survival probabilities, event times, survival curves, etc.
- Thus, we can also predict the expected difference in these values as predictor values change

Some examples

- Median survival time: find t such that $\hat{S}_i(t) = 0.5$
- The time by which $q\%$ of people with the same $\mathbf{x}_i$ have the event: find t such that $\hat{S}_i(t) = 1 - q$
- 20-week predicted survival probability: $\hat{S}_i(20)$
- Note that $\hat{S}_i(t)$ is entirely determined by the distribution used; therefore, even if multiple distributions fit the data equally well, the estimates won't be the same for each of them

R: Predicted survival quantiles

```
survprob_75_50_25 <- predict(fit, type = "quantile",
                             p = c(0.25, 0.5, 0.75),
                             se.fit = TRUE)
head(survprob_75_50_25$fit)
```

week	$\hat{S}_i(week)$	$\hat{S}_i(t) = 75\%$	$\hat{S}_i(t) = 50\%$	$\hat{S}_i(t) = 25\%$
20	0.92	46.45	86.91	142.41
17	0.83	23.17	43.35	71.03
25	0.74	24.19	45.26	74.15
52	0.87	89.76	167.95	275.19
52	0.72	46.71	87.40	143.21
52	0.71	46.05	86.17	141.19

R: Predicted (mean) event times

```
pred_time <- predict(fit, type = "response", se.fit = TRUE)
with(pred_time, head(cbind(fit, se.fit)))
```

week	pred. week	SE
20	112.84	19.84
17	56.29	8.72
25	58.76	19.10
52	218.06	67.23
52	113.48	28.89
52	111.88	19.13

R: Estimated survival probability of observed time

```
1 - psurvreg(..., mean = predict(fit, type = "lp"),
             scale = fit$scale,
             distribution = fit$dist)
```

week	$\hat{S}_i(week)$	$\hat{S}(10)$
20	0.92	0.97
17	0.83	0.92
25	0.74	0.92
52	0.87	0.99
52	0.72	0.97
52	0.71	0.97

R: Predicted change in event time

ID	$\hat{S}_i(week)$	week	new time	difference
1	0.92	20	26.26	6.26
2	0.83	17	22.32	5.32
3	0.74	25	32.82	7.82
7	0.95	23	30.19	7.19
13	0.68	37	48.57	11.57
15	0.84	25	32.82	7.82

SAS: Predicted median

```
proc lifereg data=survival.recid;
  ...
  output out=b p=med quantile=0.5 std=se cdf=cdistfunc;
run;
```

SAS: Survival probabilities

```
proc lifereg data=survival.recid outest=betahat;
  ...
  output out=b xbeta=eta;
run;
%predict(outest=betahat, out=b, xbeta=eta, time=10)
```