Parametric Survival Models

David M. Rocke

May 20, 2021

Exponential Distribution

- The exponential distribution is the base distribution for survival analysis.
- The distribution has a constant hazard λ
- The mean survival time is λ^{-1}

$$
\begin{aligned}
f(t) &= \lambda e^{-\lambda t} \\
\ln(f(t)) &= \ln\lambda - \lambda t \\
F(t) &= 1 - e^{-\lambda t} \\
S(t) &= e^{-\lambda t} \\
\ln(S(t)) &= -\lambda t \\
h(t) &= -\frac{d}{dt}\ln(S(t)) \\
&= -\frac{d}{dt}(-\lambda t) \\
&= \lambda
\end{aligned}
$$

Weibull Distribution

Using the Kalbfleisch and Prentice (2002) notation

$$
\begin{aligned}
f(t) &= \lambda p(\lambda t)^{p-1} e^{-(\lambda t)^p} \\
h(t) &= \lambda p(\lambda t)^{p-1} \\
S(t) &= e^{-(\lambda t)^p}
\end{aligned}
$$

When $p = 1$ this is the exponential. When $p > 1$ the hazard is increasing and when $p < 1$ the hazard is decreasing. This provides more flexibility than the exponential.

Exponential Regression

For each subject i define a linear predictor

$$\begin{aligned} \eta &= \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p \\ h(t|\text{covariates}) &= e^{\eta} \end{aligned}$$

We let the linear predictor have a constant term and when there are no additional predictors the hazard is $\lambda = \exp(\beta_0)$. This has a log link as in a generalized linear model. Since the hazard does not depend on t, the hazards are (trivially) proportional.

Accelerated Failure Time

Suppose that $S_i(t) = S_0(t\theta_i)$ where $\theta_i = \exp(\eta_i)$ and $\eta_i = \beta_1 x_1 + \cdots + \beta_p x_p$. This is called an accelerated failure time model because covariates cause uniform acceleration (or slowing) of failure times. If the base distribution is exponential with parameter λ then

$$S_i(t) \;=\; e^{-\lambda t\theta_i}$$

which is an exponential model with base hazard multiplied by θ_i, which is also the proportional hazards model.

Accelerated Failure Time

In terms of the log survival time $Y = \ln(T)$ the model can be written as

$$
\begin{aligned}
Y &= \alpha - \eta + W \\
\alpha &= -\ln(\lambda)
\end{aligned}
$$

where W has the extreme value distribution. The estimated parameter λ is the intercept and the other coefficients are those of η, which will be the opposite sign of those for coxph.

Accelerated Failure Time

For a Weibull distribution, the hazard function and the survival function are

$$
\begin{aligned}
h(t) &= \lambda p(\lambda t)^{p-1} \\
S(t) &= e^{-(\lambda t)^p}
\end{aligned}
$$

We can construct a proportional hazards model by using a linear predictor η_i without constant term and letting $\theta_i = e^{\eta_i}$ we have

$$
h(t) = \lambda p(\lambda t)^{p-1}\theta_i
$$

Accelerated Failure Time

A distribution with $h(t) = \lambda p(\lambda t)^{p-1}\theta_i$ is a Weibull distribution with parameters $\lambda^* = \lambda\theta_i^{1/p}$ and p so the survival function is

$$\begin{aligned} S^*(t) &= e^{-(\lambda^* t)^p} \\ &= e^{-(\lambda\theta^{1/p} t)^p} \\ &= S(t\theta^{1/p}) \end{aligned}$$

so this is also an accelerated failure time model.

Accelerated Failure Time

In terms of the log survival time $Y = \ln(T)$ the model can be written as

$$\begin{aligned} Y &= \alpha - \sigma\eta + \sigma W \\ \alpha &= -\ln(\lambda) \\ \sigma &= 1/p \end{aligned}$$

where W has the extreme value distribution. The estimated parameter λ is the intercept and the other coefficients are those of η, which will be the opposite sign of those for coxph.

Accelerated Failure Time

These AFT models are log-linear, meaning that the linear predictor has a log link. The exponential and the Weibull are the only log-linear models that are simultaneously proportional hazards models. Other parametric distributions can be used for survival regression either as a proportional hazards model or as an accelerated failure time model.

```
survreg {survival} R Documentation
Regression for a Parametric Survival Model
Description
Fit a parametric survival regression model.
These are location-scale models for an arbitrary transform of the time variable;
the most common cases use a log transformation, leading to
accelerated failure time models.

Usage
survreg(formula, data, weights, subset,
        na.action, dist="weibull", init=NULL, scale=0,
        control,parms=NULL,model=FALSE, x=FALSE,
        y=TRUE, robust=FALSE, cluster, score=FALSE, ...)

Arguments

formula
a formula expression as for other regression models. The response is usually a
survival object as returned by the Surv function.
See the documentation for Surv, lm and formula for details.

data
a data frame in which to interpret the variables named in the formula,
weights or the subset arguments.
```

```
> anderson.cox0 <- coxph(anderson.surv~treat,data=anderson)
> summary(anderson.cox0)
Call:
coxph(formula = anderson.surv ~ treat, data = anderson)

  n= 42, number of events= 30

                coef exp(coef) se(coef)     z Pr(>|z|)
treatstandard 1.5721    4.8169   0.4124 3.812 0.000138 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

              exp(coef) exp(-coef) lower .95 upper .95
treatstandard     4.817     0.2076     2.147     10.81

Concordance= 0.69  (se = 0.041 )
Likelihood ratio test= 16.35  on 1 df,   p=5e-05
Wald test            = 14.53  on 1 df,   p=1e-04
Score (logrank) test = 17.25  on 1 df,   p=3e-05
```

```
> anderson.weib <- survreg(anderson.surv~treat,data=anderson)
> summary(anderson.weib)

Call:
survreg(formula = anderson.surv ~ treat, data = anderson)
               Value Std. Error     z       p
(Intercept)    3.516      0.252 13.96 < 2e-16
treatstandard -1.267      0.311 -4.08 4.5e-05
Log(scale)    -0.312      0.147 -2.12   0.034

Scale= 0.732

Weibull distribution
Loglik(model)= -106.6   Loglik(intercept only)= -116.4
        Chisq= 19.65 on 1 degrees of freedom, p= 9.3e-06
Number of Newton-Raphson Iterations: 5
n= 42
```

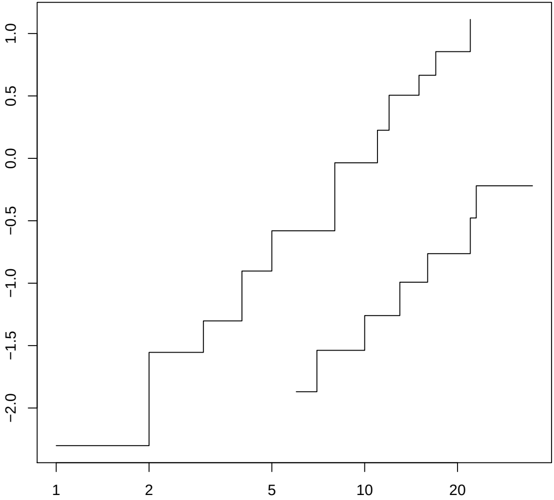
```
> anderson.exp <- survreg(anderson.surv~treat,data=anderson,dist="exp")
> summary(anderson.exp)

Call:
survreg(formula = anderson.surv ~ treat, data = anderson, dist = "exp")
              Value Std. Error     z       p
(Intercept)   3.686      0.333 11.06 < 2e-16
treatstandard -1.527     0.398 -3.83 0.00013

Scale fixed at 1

Exponential distribution
Loglik(model)= -108.5   Loglik(intercept only)= -116.8
        Chisq= 16.49 on 1 degrees of freedom, p= 4.9e-05
Number of Newton-Raphson Iterations: 4
n= 42
> plot(survfit(anderson.surv~treat,data=anderson),fun="cloglog")
```

If the cloglog plot survfit is linear, then a Weibull model may be ok.


1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
1
2
5
10
20