


Applied Survival Analysis Using R

Chapter 10: Parametric Models

Qi Guo

Department of Mathematical Sciences
The University of Texas at Dallas

April, 18 2019

- 
- 1 The Exponential Distribution
 - 2 The Weibull Model

The Exponential Distribution

- The exponential distribution is the simplest distribution to work with. It has a constant hazard function, $h(t) = \lambda$
- The p.d.f. and survival functions are, as discussed in Chap. 2, $f(t; \lambda) = \lambda e^{-\lambda t}$ and $S(t; \lambda) = e^{-\lambda t}$
- For modeling survival data, we will need the additional flexibility afforded by the Weibull distribution, of which the exponential distribution is a *special case*.

Assessing the Weibull Distribution as a Model for Survival Data in a Single Sample

- The Weibull survival distribution, as expressed in Chapter 2, as hazard and survival functions $h(t) = \alpha \lambda^\alpha t^{\alpha-1}$ and $S(t) = e^{-(\lambda t)^\alpha}$
- Use the scale parameter $\sigma = 1/\alpha$ with parameterization $\mu = -\log \lambda$

$$h(t) = \frac{1}{\sigma} e^{-\frac{\mu}{\sigma}} t^{\frac{1}{\sigma}-1} \quad (1)$$

and

$$S(t) = e^{-e^{-\mu/\sigma} t^{1/\sigma}} \quad (2)$$

- Taking a complementary log-log transformation $g(\mu) = \log[-\log(\mu)]$ of the Weibull survival function

Assess how well a set of survival data

$$\log[-\log(S_i)] = \alpha \log(\lambda) + \alpha \log(t_i) = -\frac{\mu}{\sigma} + \frac{1}{\sigma} \log(t_i) \quad (3)$$

- where $S_i = S(t_i)$
- Assess how well a set of survival data follow a Weibull distribution:
 - First compute the Kaplan-Meier estimate \hat{S} of a survival distribution.
 - Define $y_i = \log \{ -\log[\hat{S}(t_i)] \}$ and plot y_i versus $\log(t_i)$.
 - Fit through these points a straight line, with equation of the form $y = b + m \log t$ where $b = -\mu/\sigma$ and $m = 1/\sigma$.

Example

- Examine the `gastricXelox` data discussed in Chapter 3 to see if it follows a Weibull distribution.
- First obtain a Kaplan-Meier estimate of the survival distribution

```
1 > timeMonths <- gastricXelox$timeWeeks*7/30.25
2 > delta <- gastricXelox$delta
3 > library(survival)
4 > result.km <- survfit(Surv(timeMonths, delta) ~ 1)
```

- Extract the survival estimates and time variables from “`result.km`” and transform the former with a complementary log-log transformation, and the latter with a log transformation

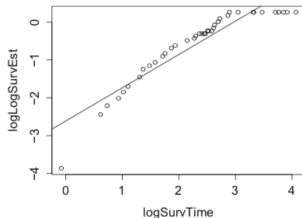
```
1 > survEst <- result.km$surv
2 > survTime <- result.km$time
3 > logLogSurvEst <- log(-log(survEst))
4 > logSurvTime <- log(survTime)
```

Plot

- Finally, we plot “logLogSurvEst” versus “logSurvTime” and fit a straight line through the points

```
1 > plot(logLogSurvEst ~ logSurvTime)
2 > result.lm <- lm(logLogSurvEst ~ logSurvTime)
3 > abline(result.lm)
```

Fig. 10.1 Plot of the complementary log-log transformation of survival probability versus log survival time for the gastricXelox data



- Conclusion: The points do not follow a linear relationship. Indicate that a Weibull distribution is not appropriate.

Maximum Likelihood Estimation of Weibull Parameters for a Single Group of Survival Data

- The log-likelihood function in Chapter 2,

$$\ell(\lambda, \alpha) = \sum_{i=1}^n \{\sigma_i \log[h(t_i)] + \log[S(t_i)]\} \quad (4)$$

- Substituting the expressions for $h(t_i)$ and $S(t_i)$

$$\ell(\lambda, \alpha) = d \log \alpha + d\alpha \log \lambda + (\alpha - 1) \sum_{i=1}^n \sigma_i \log t_i - \lambda^\alpha \sum_{i=1}^n t_i^\alpha \quad (5)$$

where $d = \sum_{i=1}^n \delta_i$

- Re-parametrize in terms of $\alpha = 1/\sigma$ and $\lambda = e^{-\mu}$

In R

```

1 logLikWeib <- function(par, tt, status) { mu <- par[1]
2 sigma <- par[2]
3 lambda.p <- exp(-mu)
4 alpha.p <- 1/sigma
5 dd <- sum(status)
6 sum.t <- sum(status*log(tt))
7 sum.t.alpha <- sum(tt^alpha.p)
8 term.1 <- dd*log(alpha.p) + alpha.p*dd*log(lambda.p)
9 term.2 <- (alpha.p - 1)*sum.t
10 term.3 <- (lambda.p^alpha.p)*sum.t.alpha
11 result <- term.1 + term.2 - term.3
12 result }

```

- The m.l.e may be obtained using the “optim” function, using as starting values the estimates of μ and σ from the linear regression

```

1 result <- optim(par=c(4.568, 2.280), fn=logLikWeib,
2 method= "L-BFGS-B", lower=c(0.001, 0.01), upper=c(5, 5),
3 control=list(fnscale = -1), tt=ttr, status=relapse)

```

- Option “control=list(fnscale = -1)” to tell the optim function to find a maximum (rather than a minimum).

In R

- The final m.l.e. is given by

```
1 > result$par
2 [1] 4.656329 2.041061
```

The first element of “`result$par`” is $\hat{\mu}$ and the second element is $\hat{\sigma}$

- Use function in R:

```
1 > result.survreg.0 <- survreg(Surv(ttr, relapse) ~ 1,
2 dist="weibull")
3 > summary(result.survreg.0)
```

	Value	Std. Error	z	p
(Intercept)	4.656	0.2170	21.46	3.68e-102
Log(scale)	0.713	0.0919	7.76	8.26e-15

```
7
8 Scale= 2.04
```

Selecting a Weibull Distribution to Model Survival Data

- Suppose the two time points are t_1 and t_2 , and the estimated survival points (from the Kaplan-Meier survival curve) at these two points are s_1 and s_2 , define $y_1 = \log[-\log(s_1)]$ and $y_2 = \log[-\log(s_2)]$, using (3).

$$y_1 = \alpha \log(\lambda) + \alpha \log(t_1) \quad y_2 = \alpha \log(\lambda) + \alpha \log(t_2) \quad (6)$$

- Solving these two simultaneous linear equations, we get

$$\tilde{\alpha} = \frac{y_1 - y_2}{\log(t_1) - \log(t_2)} \quad \tilde{\lambda} = \exp\left\{\frac{y_2 \log(t_1) - y_1 \log(t_2)}{y_1 - y_2 - 2}\right\} \quad (7)$$

Example

- Find a Weibull distribution that matches the Kaplan-Meier estimate of the survival distribution at 28 and 84 days.

```

1 > result.surv <- survfit(Surv(ttr, relapse) ~ 1,
2 subset=grp == "patchOnly")
3 > result.summ <- summary(result.surv, time=c(28, 84))
4 > t.vec <- result.summ$time
5 > s.vec <- result.summ$surv
6 > data.frame(t.vec, s.vec)
7   t.vec   s.vec
8 1    28 0.437500
9 2    84 0.265625

```

- Use the “Weibull2” function in “Hmisc” package to produce a Weibull function that matches these two points.

```

1 > library(Hmisc)
2 > pharmWeib <- Weibull2(t.vec, s.vec)

```

Example

- The function “`pharmWeib`” computes the Weibull survival estimates for a range of time values.

```
1 > t.vals <- 1:200
2 > s.vals <- pharmWeib(t.vals)
```

- Obtain the predicted Weibull survival curve based on maximum likelihood estimates of the Weibull parameters.

```
1 > model.pharm.weib.basic <- survreg(Surv(ttr, relapse) ~ 1,
2 dist="weibull", subset=grp == "patchOnly" )
3 > mu.hat <- model.pharm.weib.basic$coefficients
4 > sigma.hat <- model.pharm.weib.basic$scale
5 > lambda.hat <- exp(-mu.hat) # "1 / scale"
6 > alpha.hat <- 1/sigma.hat # "shape"
7 > s.mle.vals <- 1 - pweibull(t.vals, shape=alpha.hat,
8 scale=1/lambda.hat)
```

Plot

- Finally, plot the survival estimates in Fig. 10.4

```

1 > plot(result.surv, conf.int=F, xlab="Days to relapse", ylab="Survival probability")
2 > lines(s.mle.vals ~ t.vals, col="blue")
3 > lines(s.vals ~ t.vals, col="red")
4 > points(t.vec, s.vec, col="red")

```

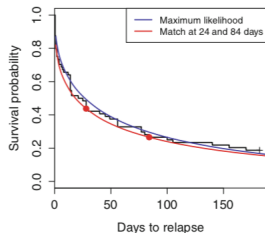


Fig. 10.4 Survival curve estimates for the “patch only” group in the pharmacoSmoking data. The step function is the Kaplan-Meier estimate. The *blue line* is the Weibull estimate of the survival curve based on maximum likelihood estimates of the parameters. The *red line* is the Weibull estimate that matches the Kaplan-Meier estimate at 24 and 84 days; the two matching points are indicated by *solid red circles*

Comparing Two Weibull Distributions Using the Accelerated Failure Time and Proportional Hazards Models

- Comparing a treatment group to a control group, often used with parametric models, is called *the accelerated failure time (AFT) model*.
- Assume that the survival time for a treated patient is a multiple e^γ of what the survival time would have been had the patient received the control treatment.
- If the treatment is effective, the accelerated time coefficient e^γ will be *greater* than one, and thus γ will be positive.
- Formally, the survival distributions for the accelerated life model are given by $S_1(t) = S_0(e^{-\gamma}t)$ and the hazards are given by $h_1(t) = e^{-\gamma}h_0(e^{-\gamma}t)$

Example

- In the case of the Weibull distribution.

$$h_1(t) = e^{-\frac{\gamma}{\sigma}} \cdot \frac{1}{\sigma} \cdot e^{-\frac{\mu_0}{\sigma}} t^{\frac{1}{\sigma}-1} = e^{-\frac{\gamma}{\sigma}} h_0(t) \quad (8)$$

- The pharmacoSmoking data, comparing the triple therapy treatment group to the patch treatment.

```

1 > result.survreg.grp <- survreg(Surv(ttr, relapse) ~ grp,
2   dist="weibull")
3 > summary(result.survreg.grp)
4           Value      Std. Error      z      p
5 (Intercept)   5.286      0.3320    15.92  4.59e-57
6 grppatchOnly -1.251      0.4348     -2.88  4.00e-03
7 Log(scale)    0.689      0.0911     7.56  3.97e-14
8
9 Scale= 1.99

```

- Compare the patch group to the triple therapy group using a proportional hazards model, the log proportional hazards is given by $\hat{\beta}_a = -\hat{\gamma}/\hat{\sigma} = 1.251/1.99 = 0.629$

Example

- Compare this to the results of fitting a Cox proportional hazards model as follows

```

1 > result.coxph.grp <- coxph(Surv(ttr, relapse) ~ grp)
2 > summary(result.coxph.grp)
3   n= 125, number of events= 89
4           coef      exp(coef)    se(coef)      z    Pr(>|z|)
5 grppatchOnly  0.6050      1.8313    0.2161    2.8  0.00511 **

```

- Obtain the baseline Weibull coefficient survival function in R as follows:

```

1 mu0.hat <- result.survreg.grp$coef[1]
2 sigma.hat <- result.survreg.grp$scale
3 alpha.hat <- 1/sigma
4 lambda0.hat <- exp(-mu0.hat)
5 tt.vec <- 0:182
6 surv0.vec <- 1 - pweibull(tt.vec, shape=alpha, scale=1/lambda0.hat)

```

- In R, $\hat{\gamma}$ is the coefficient for the “grp” term, and is the second element of “coef”,

Example

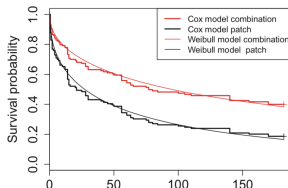
- compare these survival estimates to those from the Cox proportional hazards model

```
1 > coxph.surv.est <- survfit(result.coxph.grp,
2 > newdata=data.frame(list(grp=c("combination", "patchOnly"))))
```

- plot the Cox-based survival curves and the Weibull-based survival curves on the same plot

```
1 > plot(coxph.surv.est, col=c("red", "black"))
2 > lines(surv0.vec ~ tt.vec, col="red")
3 > lines(surv1.vec ~ tt.vec)
```

Fig. 10.5 Comparisons of combination therapy (*red*) vs. patch (*black*) for time to smoking relapse using the *pharmacSmoking* data. The step functions are survival function estimates obtained using a Cox proportional hazards model, and the *smooth curves* are obtained using a Weibull model



Model Selection and Residual Analysis with Weibull Survival Data

- Fit a model with all covariates as predictors, and then use backwards stepwise regression, using the AIC as a measure of goodness of fit, as follows:

```
1 > modelAll.pharm.weib <- survreg(Surv(ttr, relapse) ~ grp +
2 wogender + race + employment + yearsSmoking + levelSmoking +
3 age + priorAttempts + longestNoSmoke, dist="weibull")
4 > model.step.pharm.weib <- step(modelAll.pharm.weib)
```

- Use the “residuals” function to compute deviance residuals and deletion residuals.

```
1 resid.deviance <- residuals(model.pharm.weib, type="deviance")
2 par(mfrow=c(2,2))
3 plot(resid.deviance ~ age)
4 smoothSEcurve(resid.deviance, age)
5 title("Deviance residuals versus age")
6 plot(resid.deviance ~ grp)
7 title("Deviance residuals versus treatment group")
8 plot(resid.deviance ~ employment)
```

Plot

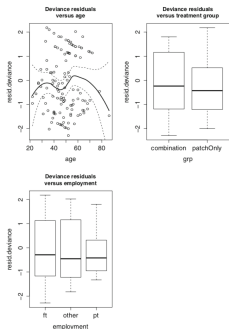


Fig. 10.6 Deviance residual plots from Weibull model fit to the pharmacSmoking data

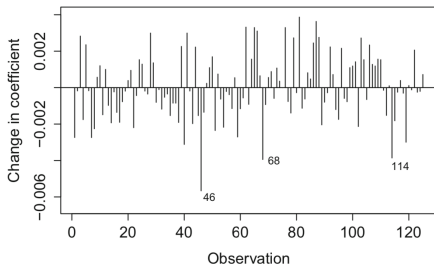
- The effects of individual patients on the estimate of the coefficient for “age”

Plot

```

1 resid.dfbeta <- residuals(model.pharm.weib, type="dfbeta")
2 n.obs <- length(ttr)
3 index.obs <- 1:n.obs
4 plot(resid.dfbeta[,3] ~ index.obs, type="h",
5      xlab="Observation", ylab="Change in coefficient",
6      ylim=c(-0.0065, 0.004)) abline(h=0)

```



- We see that patients 46 and 68 are again influential, as is patient 114.