Department of Biostatistics and Bioinformatics
EMORY ROLLINS SCHOOL OF PUBLIC HEALTH

BIOS 522: Survival Analysis Methods

Reading 9:
Parametric regression models

This week, we will define the parametric proportional hazards model and accelerated failure time model.

Part 1. Parametric proportional hazards model

Parametric proportional hazards model

The Cox proportional hazards regression model is a semiparametric regression model, where the baseline hazard function is left flexible, and the effect of the covariates is modeled through the parametric hazard ratio term. If we were to set the baseline hazard function to a known parametric distribution (e.g. exponential, Weibull), the new model is fully parametric. This is known as a **parametric proportional hazards model**.

Features of parametric survival models:

- (+) More efficient when the data follow a parametric distribution
- (+) Produce smoothed distributions, even with small/sparse data sets
- (+) Readily generate any prediction (mean, median, quantile)
- (+) Well suited to study the shape of the hazard function, not just the hazard ratio, because it produces a smooth hazard function
- (+) Can be generalized to model complex settings, like jointly modeling several correlated survival outcomes, or interval censored data (*next week's topic…*)
- (-) May incur bias if model is misspecified – need to justify your assumptions!

Weibull proportional hazards models

In a Weibull proportional hazards model, we model the effect of the covariates using the standard proportional hazards approach, but set the baseline hazard function to $h_0(t) = \lambda_0\gamma_0(\lambda_0 t)^{\gamma_0-1}$, corresponding to the hazard function for a Weibull distribution with rate parameter λ_0 and shape parameter γ_0, e.g.:

$$h_i(t) = \lambda_0\gamma_0(\lambda_0 t)^{\gamma_0-1}\exp(\beta_1 X_{i1} + \cdots + \beta_1 X_{ik})$$

Thus, the reference group has the hazard function of a Weibull random variable with rate λ_0 and shape γ_0, and $\exp(\beta_1 X_{i1} + \cdots + \beta_1 X_{ik})$ measures the multiplicative impact of the covariates on the hazard function for individuals not in the reference group. As in the Cox model, $\exp(\beta_j)$ is the hazard ratio for a one-unit increase in covariate X_{ij}, holding all other covariates constant.

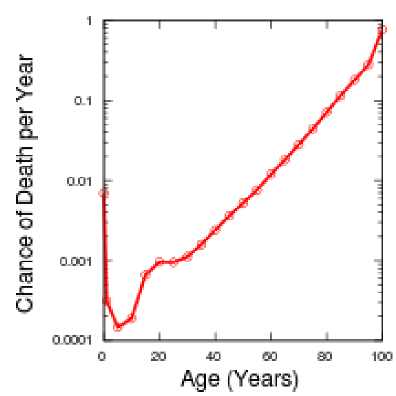
The Weibull proportional hazards model can be fit by maximum likelihood, where the hazard and survival functions implied by the above model are plugged into the likelihood function for each individual i:

$$L(\lambda_0, \gamma_0, \beta_1, \dots, \beta_k) = \prod_{i=1}^{n} [h(T_i^*|\lambda_0, \gamma_0, \beta_1, \dots, \beta_k)]^{\delta_i} S(T_i^*|\lambda_0, \gamma_0, \beta_1, \dots, \beta_k)$$

Maximizing this likelihood gives us the MLEs for the baseline rate and shape parameters and the coefficients $\beta_1, \dots, \beta_k$.

Piecewise exponential models

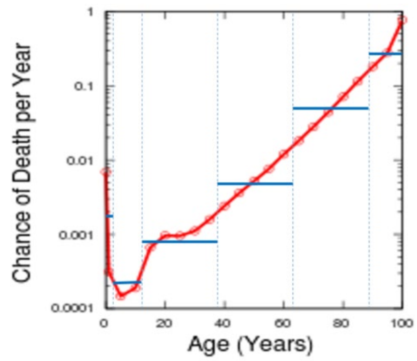
Importantly, parametric distributions do not accommodate all possible hazard function shapes. Consider, for example, human mortality (right).



This shape cannot be modeled by a Weibull or log-logistic hazard function. In this case, we would likely use a Cox proportional hazards model, but there may be settings where we prefer a parametric approach.

An alternative parametric approach with more flexibility is to model the baseline hazard function with a piecewise constant hazard function with breakpoints determined by the user. Because the exponential assumption assumes constant hazard, this is often referred to as a **piecewise exponential model**.

Consider fitting six-time intervals to model human mortality. The piecewise constant hazard in each interval is shown in blue:



More generally, we can divide time into J non-overlapping intervals $[0, \tau_1), [\tau_1, \tau_2), \ldots, [\tau_{J-1}, \infty)$. The cut points $\tau_1, \ldots, \tau_{J-1}$ are defined by the user. Note that intervals do not need to be equally sized.

Since we assume that the baseline hazard is constant within each time interval, the baseline hazard function $h_0(t)$ can be expressed as:

$$h_0(t) = \begin{cases} \lambda_1 & 0 \le t < \tau_1 \\ \lambda_2 & \tau_1 \le t < \tau_2 \\ \ldots & \\ \lambda_J & t \ge \tau_{J-1} \end{cases}$$

Thus, we model the baseline hazard $h_0(t)$ using J parameters $\lambda_1, \ldots, \lambda_J$, each representing the hazard rate for the reference group in one particular interval.

We can then model the effect of covariates using a proportional hazards model:

$$h_i(t) = h_0(t) \exp(\beta_1 X_{i1} + \cdots + \beta_k X_{ik})$$

The coefficients are log hazard ratios for a one-unit increase in each covariate, adjusting for other covariates. We assume proportional hazards for each covariate. Though we flexibly model the baseline hazard function using piecewise constant intervals, the above model assumes that the effect of the covariate on the hazard is constant over time, i.e. same β_j for all time intervals.

The piecewise constant approach can accommodate a wide range of hazard functions. A disadvantage is that the user must define the number of intervals and the location of the cut points subjectively. In the absence of external information, a common strategy is to select intervals with equal numbers of events in each. Lindsey and Ryan (Statistics in Medicine, 1998) suggest picking intervals so that there are approximately 10 events per interval. The change points do not need to be equidistant.

Part 2. Accelerated failure time model

Introduction to the AFT model

The **accelerated failure time (AFT) regression model** is another class of parametric regression models for survival data.

In the *proportional hazards model*, the dependent variable is the *hazard function $h_i(t)$ for individual i*. The hazard function in the reference group is $h_0(t)$. The covariates multiply the baseline hazard function via the *hazard ratio*.

$$h_i(t) = h_0(t)\exp(\beta_1 X_{i1} + \cdots + \beta_k X_{ik})$$

In the *accelerated failure time model*, the dependent variable is the *survival time random variable T_i for individual i.* The reference group population has survival time random variable T_0. The impact of the covariates is to speed up or slow down time in the following fashion:

$$T_i = \exp(\alpha_1 X_{i1} + \cdots + \alpha_k X_{ik})T_0$$

Consider a simplified model with only a single binary covariate X_i. Then the group with $X_i = 1$ has survival times T_1 expressed as the reference group random variable times the acceleration factor:

$$T_1 = \exp(\alpha)\, T_0$$

We refer to α as the **coefficient** and $\exp(\alpha)$ as the **acceleration factor**. If the acceleration factor is 2, group 1 survives on average twice as long as group 0. The acceleration factor models the effect of the covariate X_i.

An equivalent way to express the accelerated failure time model is by each group's survival function. Consider that the survival function in the reference group is $S_0(t)$. For individual i, the random variable T_i has survival $S_i(t)$, which can be expressed as a function of the reference group survival and the coefficients.[1]

$$S_i(t) = S_0\left(e^{-(\alpha_1 X_{i1} + \cdots + \alpha_k X_{ik})}t\right)$$

The impact on the time scale is measured by the term $e^{-(\alpha_1 X_{i1} + \cdots + \alpha_k X_{ik})}$.

Like the parametric proportional hazards models introduced in the previous section, we assume that survival times in the reference/baseline group follow

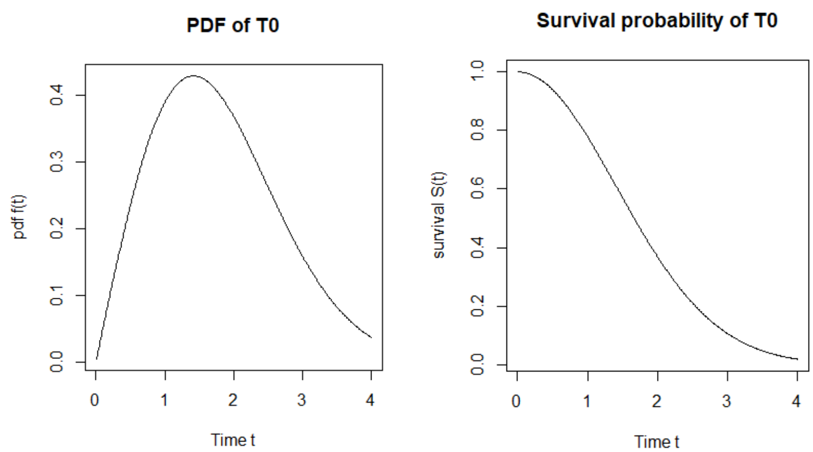
[1] *Note*: For those interested in seeing how one model format directly implies the other:

$$S_1(t) = \Pr(T_1 > t) = \Pr(\exp(\alpha)\, T_0 > t) = \Pr\left(T_0 > \frac{t}{\exp(\alpha)}\right) = S_0(\exp(-\alpha)\, t)$$

a known parametric distribution (e.g. exponential, Weibull, log-logistic). Thus, the accelerated failure time model is a fully **parametric survival model**.

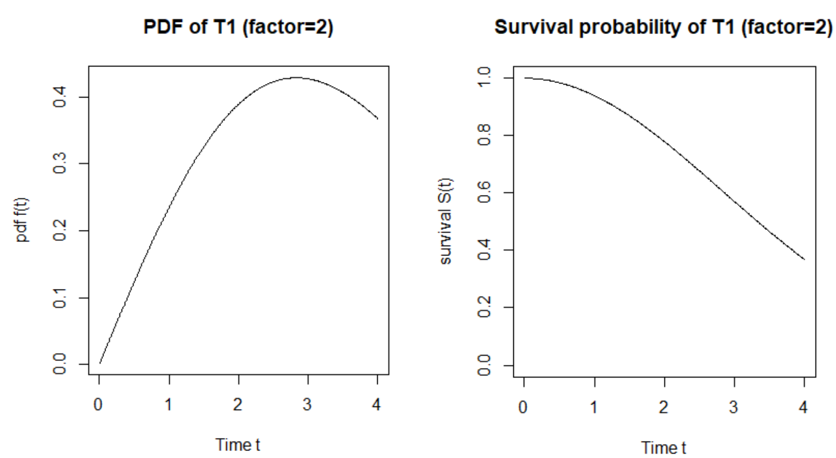
Stretching time

We use a simplified example with a single covariate to demonstrate the effect of the acceleration factor on the pdf and survival function. Imagine that the reference group has survival times T_0 that follow a Weibull distribution with rate $\lambda_0 = 0.5$ days^{-1} and shape $\gamma = 2$. The pdf $f(t)$ of T_0 is plotted on the left. The survival function $S(t)$ of T_0 is plotted on the right.



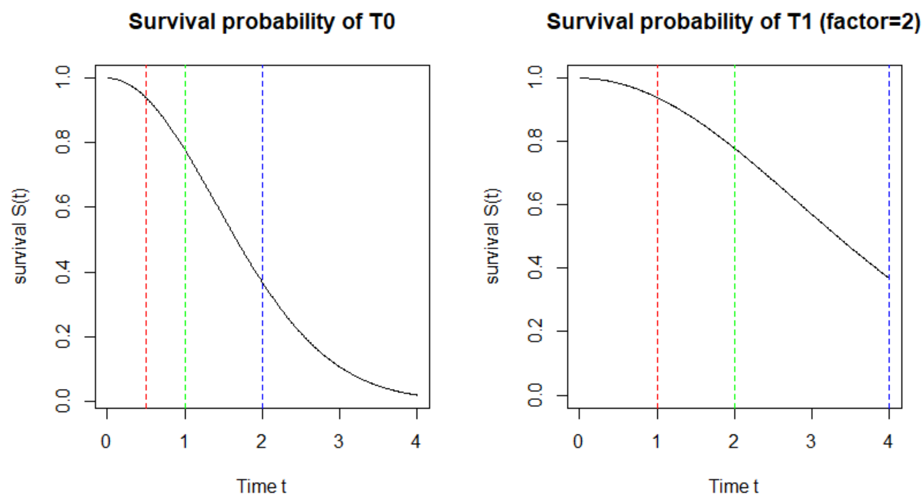
We see that most failure times are between 1 and 2 days. Median survival is 1.67 days.

Now imagine that our coefficient $\alpha = 0.693$. The acceleration factor is $\exp(\alpha) = \exp(0.693) = 2$, and $T_1 = 2T_0$. Group 1 has survival times that are twice as long as the survival times for Group 0. The pdf $f(t)$ of T_1 is plotted on the left for an acceleration factor of 2. The survival function $S(t)$ of T_1 is plotted on the right for an acceleration factor of 2.



Most failure times in this group are between 2 and 4 days. Median survival is 3.33 days, which is twice as long as group 0.

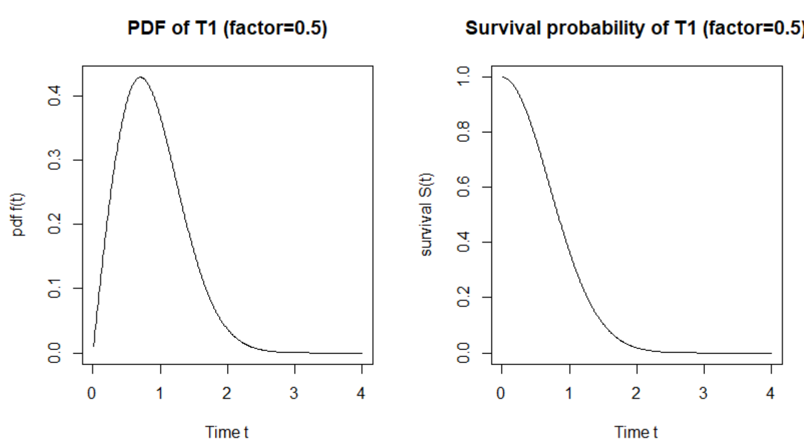
The survival functions for the two groups are closely related. The curve for the group with $X_i = 1$ is a “stretched out” version of the curve for the group with $X_i = 0$. They are plotted again below with helpful vertical lines.



Consider the red dashed line. In the group with $X_i = 0$, the survival probability at time $t = 0.5$ is equal to the survival probability in the group with $X_i = 1$ at time $t = 1$. In other words, $S_1(1) = S_0(0.5)$. The green and blue dashed lines show similar relationships.

In this way, the acceleration factor has expanded (slowed down) time in the group with $X_i = 1$. Their survival times are twice as long. These subjects effectively “age” at only half the normal speed.

If instead our coefficient is $\alpha = -0.693$, then the acceleration factor is $\exp(\alpha) = \exp(-0.693) = 0.5$, and $T_1 = 0.5T_0$. Group 1 has survival times that are half as long as the survival times for Group 0. The pdf $f(t)$ of T_1 is plotted on the left. The survival function $S(t)$ of T_1 is plotted on the right.



Most failure times in this group are between 0.5 and 1 day. Median survival is 0.833 days, which is half as long as the other group. The acceleration factor has compressed (sped up) time in the group with $X_i = 1$. These subjects effectively "age" at twice the normal speed.

We can summarize the relationship between the coefficient, the acceleration factor, and survival in the following table.

Coefficient	**Acceleration factor**	**Survival in group 1 versus group 0**
$\alpha > 0$	$\exp(\alpha) > 1$	Increased survival; ages slower
$\alpha = 0$	$\exp(\alpha) = 1$	No difference in survival
$\alpha < 0$	$\exp(\alpha) < 1$	Decreased survival; ages faster

More generally, we can have several different covariates in our model.

$$T_i = \exp(\alpha_1 X_{i1} + \cdots + \alpha_k X_{ik}) T_0$$

We combine the impact of each covariate. Some may speed up time, and some may slow down time. For an individual with covariates $X_{i1}, \ldots, X_{ik}$, the term $\exp(\alpha_1 X_{i1} + \cdots + \alpha_k X_{ik})$ measures how much longer on average that individual will survive than someone from the reference group.

Fitting an AFT model

To fit an AFT model, we first must pre-specify the parametric distribution we will use (e.g. exponential, Weibull, log-logistic) and the covariates we will include. If we have k covariates, we will need to fit k model coefficients $(\hat{\alpha}_1, \ldots, \hat{\alpha}_k)$. In addition, we need to estimate the parameters that define the reference group distribution. For an exponential distribution, we need to fit the rate parameter $\hat{\lambda}_0$ in the reference group. For a Weibull or log-logistic distribution, we need to fit the rate parameter $\hat{\lambda}_0$ and shape parameter $\hat{\gamma}_0$ in the reference group.

Given the data, we can use maximum likelihood methods to estimate the $k + 1$ or $k + 2$ total parameters in the model. We obtain our maximum likelihood estimates of the reference group parameters $\hat{\lambda}_0$ and $\hat{\gamma}_0$ (or just $\hat{\lambda}_0$ if fitting an exponential AFT model). We also obtain maximum likelihood estimates of the coefficients $\hat{\alpha}_1, \ldots, \hat{\alpha}_k$.

We are rarely interested in reporting the coefficients directly. We instead prefer to report the acceleration factor for each covariate. To calculate the acceleration ratio, we exponentiate our coefficient, e.g. $\exp(\hat{\alpha}_1)$.

Comparing the AFT to other survival models

Importantly, even though both AFT and parametric proportional hazards models are parametric survival models, they are not in general equivalent. A log-logistic proportional hazards model is not equal to a log-logistic AFT model.

In certain settings, though, the models overlap. One can show that the exponential proportional hazards model is equivalent to the exponential AFT model. Similarly, the Weibull proportional hazards model is equivalent to the Weibull AFT model. (See homework).

Nonetheless, as noted earlier, acceleration factors are not interpreted in the same way as hazard ratios, and so it is important to pay attention to which model is actually fit when interpreting the results. But the conclusions will be the same when the models are equivalent. Even when the models are slightly different (e.g. Cox model vs. Weibull AFT), the *general* conclusions will tend to be similar unless one model fits very poorly.

Part 3. Looking ahead

Parametric models are particularly useful for certain complex data types. Next week, we will introduce two important concepts. The first is left truncation, when individuals enter into a study later than time 0. The second is interval censoring, when failure times are not directly observed but rather only known to lie within an interval.