The Weibull Distribution

Patrick Breheny

October 9

Introduction

- Today we will introduce an important generalization of the exponential distribution called the Weibull distribution
- Unlike the exponential distribution, in which hazards are restricted to be constant, the Weibull distribution allows hazards to increase or decrease over time
- In this lecture, we will derive the Weibull distribution, explore various connections between the Weibull and other distributions we have seen, and finally, introduce Weibull regression

Motivation

- Suppose $X \sim \text{Exp}(\tau)$, and consider the transformation $T = X^\sigma$
- The resulting distribution is called the *Weibull distribution* and its hazard function is given by

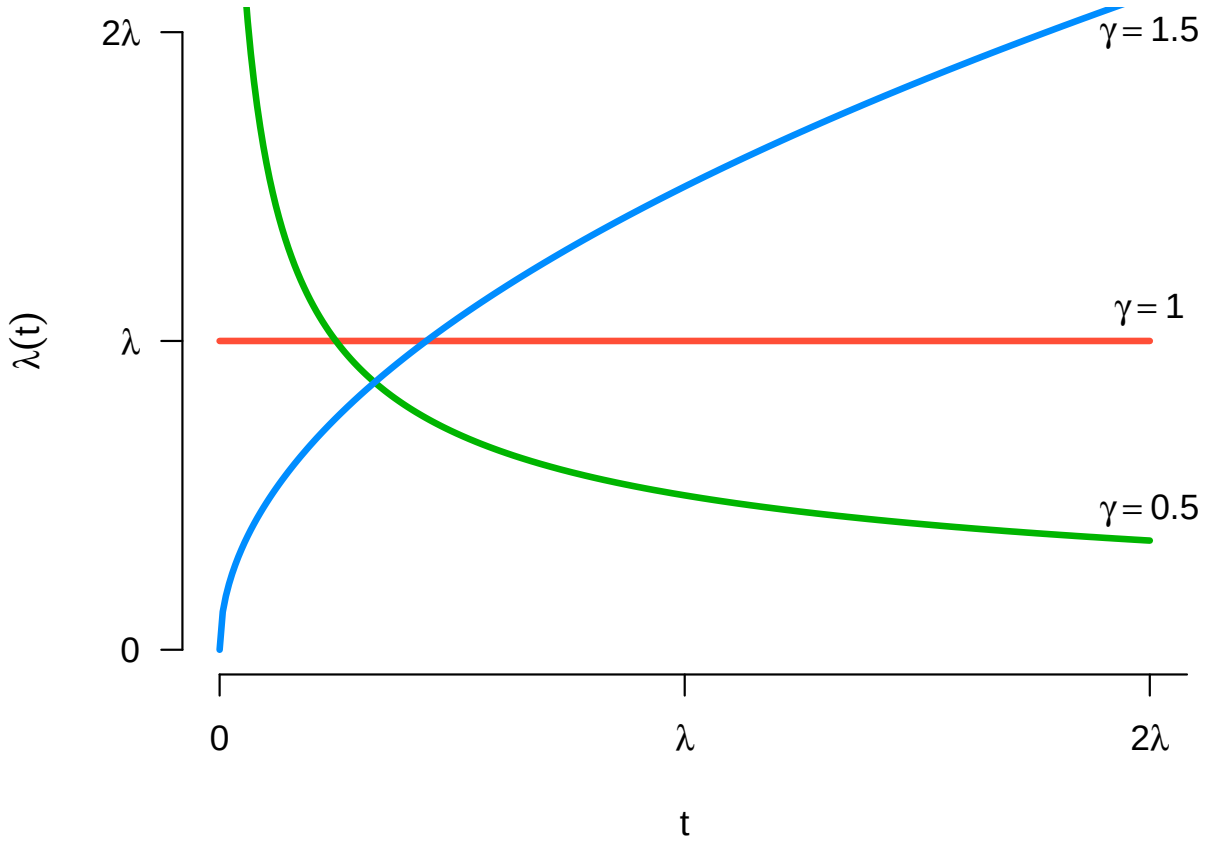
$$\lambda(t) = \frac{\tau}{\sigma} t^{1/\sigma - 1}$$

- This can be reparameterized in various ways:
  - Our book uses the parameterization

$$\lambda(t) = \lambda\gamma(t\lambda)^{\gamma-1},$$

    where $\gamma = 1/\sigma$ and $\lambda^\gamma = \tau$
  - The `R` functions `dweibull`, `pweibull`, etc., use the same parameterization except in terms of a scale parameter $\beta = 1/\lambda$ instead of a rate parameter

Illustration: Various Weibull distributions



Remarks

- Thus, by setting $\gamma > 1$, the Weibull distribution can accommodate *positive aging*, in which the risk of failure goes up over time
- Example: The Pike rate data, in which no deaths occurred in the first 142 days, but 32 of the 36 deaths occurred in the next 142 days; positive aging is present here because it takes time for the cancer to progress to the point where it is fatal
- By setting $\gamma < 1$, the Weibull distribution can accommodate *negative aging*, in which the risk of failure goes down over time
- Example: The GVHD data, in which all 20 events occur in the first 49 days and no events occurred in the next 1,308 days; negative aging is present here because GVHD is a complication of marrow transplantation

Remarks (cont'd)

- Note that, unlike the Gamma distribution, this positive and negative aging is not bounded:
  - When $\gamma > 1$, $\lambda(t) \to \infty$ as $t \to \infty$
  - When $\gamma < 1$, $\lambda(t) \to 0$ as $t \to \infty$
- With various combinations of λ and γ, the Weibull distribution is quite flexible at representing various time-to-event processes
- However, it is still unable to represent hazards that both rise and fall with time, such as human mortality from birth

Diagnostic plot

- Note that the Weibull distribution has cumulative hazard and survival functions

$$\Lambda(t) = (\lambda t)^\gamma$$
$$S(t) = \exp\{-(\lambda t)^\gamma\}$$

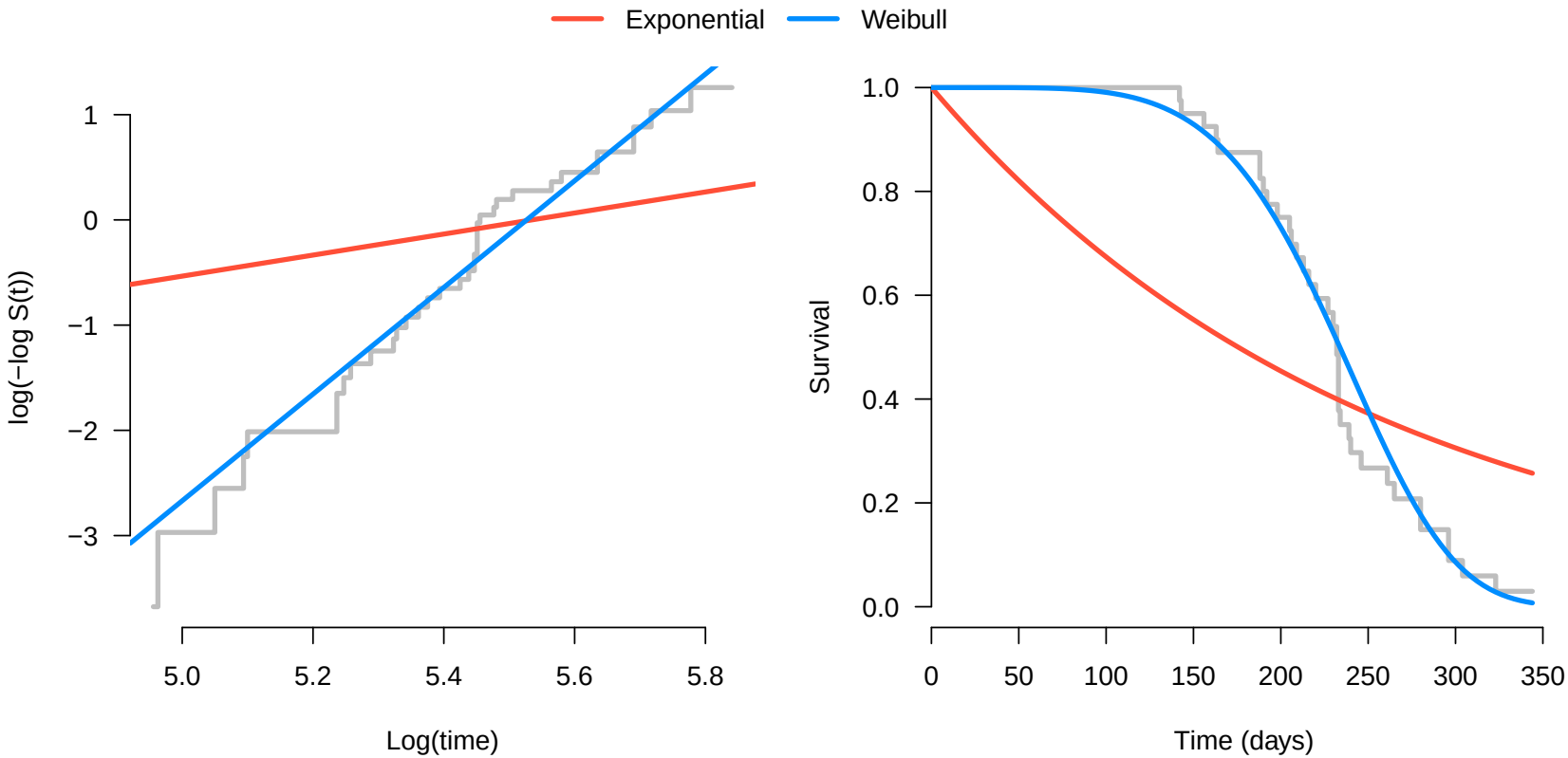
- This suggests the diagnostic plot

$$\log\{-\log \hat{S}(t)\} = \gamma(\log \lambda + \log t);$$

in other words, a plot of the complimentary log-log of the Kaplan-Meier estimate against $\log t$ should be linear

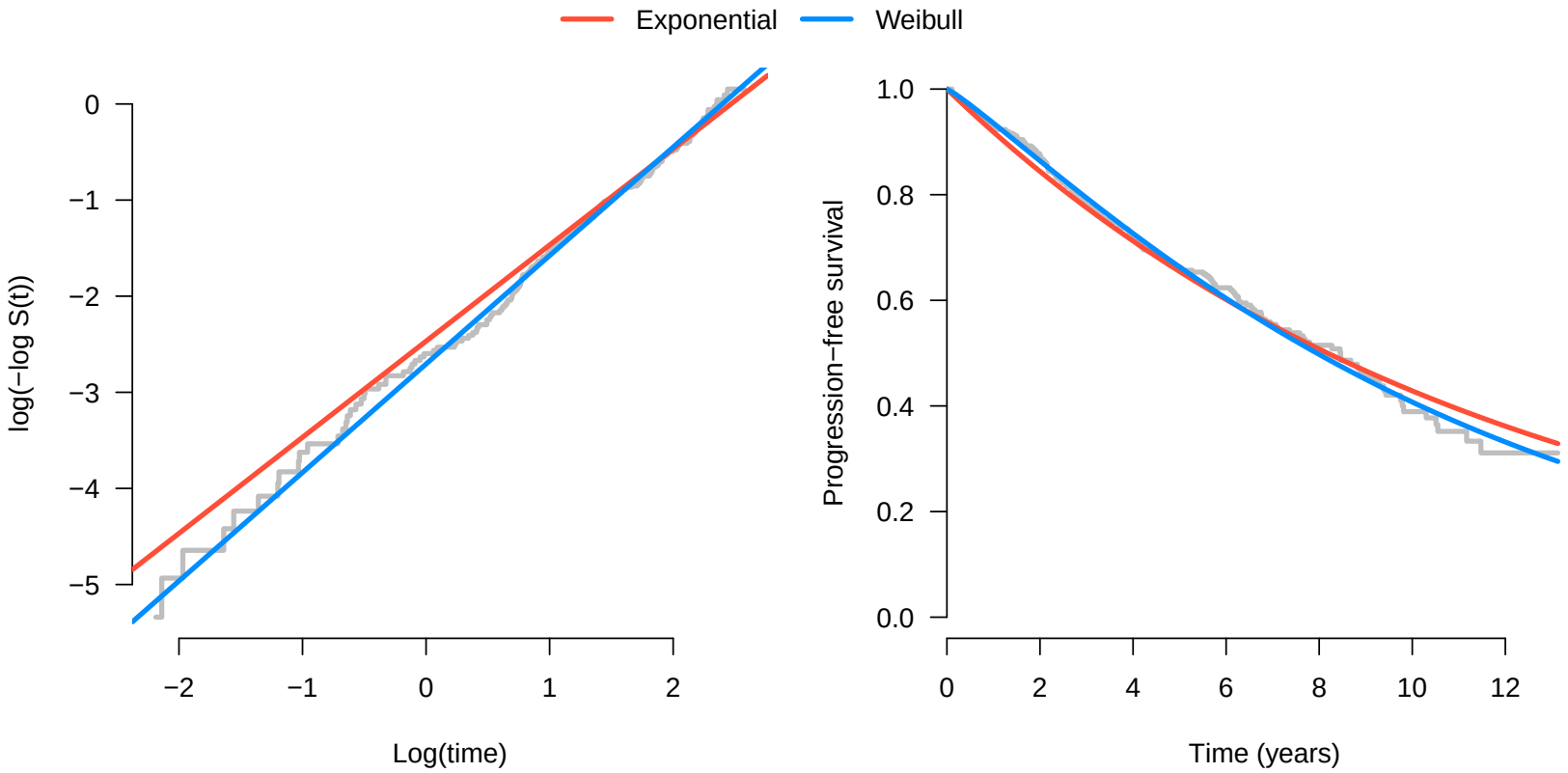
Pike rat data

Weibull vastly superior to exponential



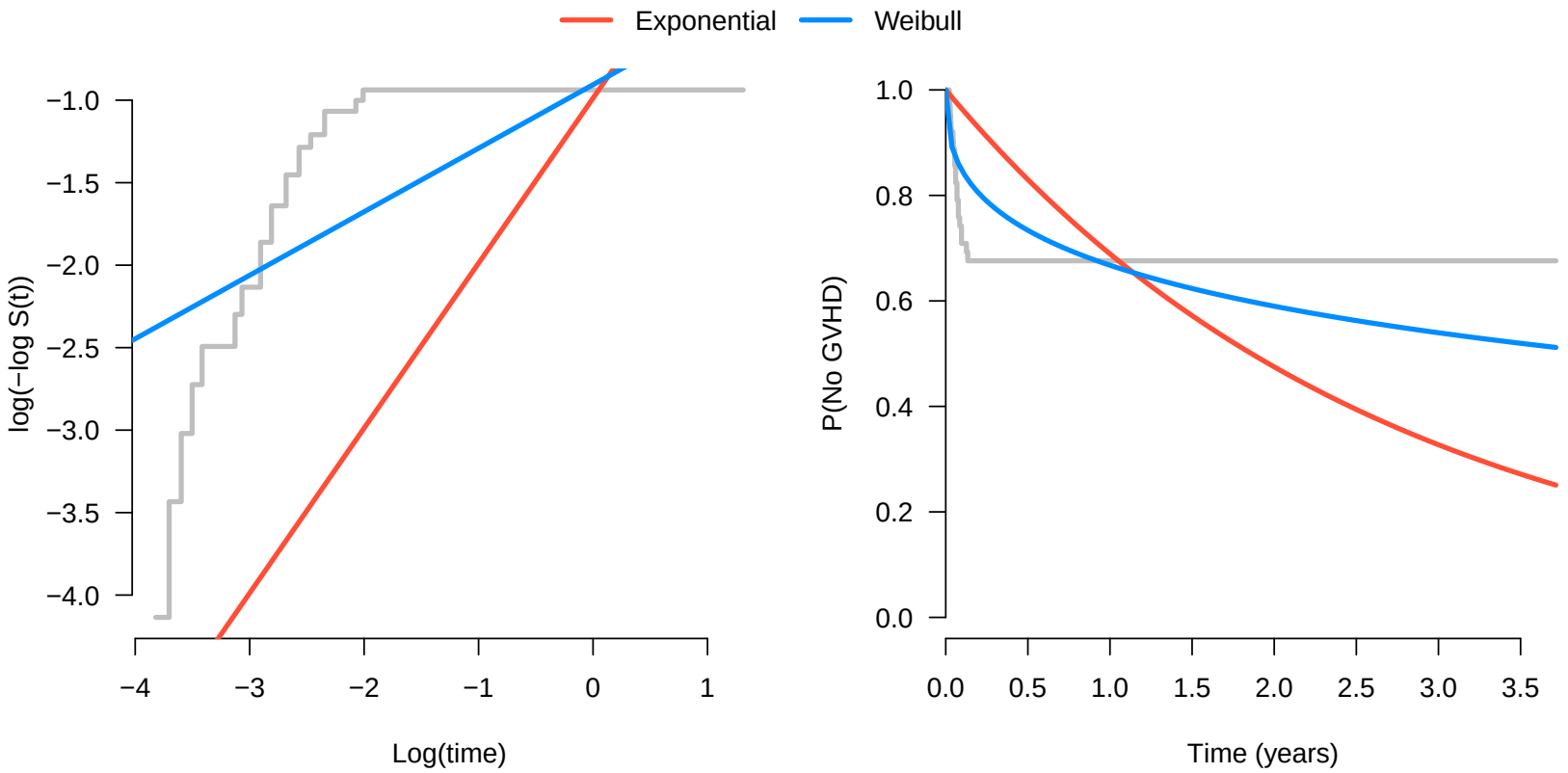
PBC data

Both Weibull and exponential seem fine



GVHD data

Neither Weibull nor exponential is very good



Remarks

- Occasionally, survival distributions really do follow what look like exponential distributions
- There are many occasions, however, in which the Weibull distribution is a much more accurate model of failure time process than the exponential
- However, there are also distributions that don't resemble any parametric models, such as events which may never occur (sometimes referred to as "defective" distributions)

The standard extreme value distribution

- Suppose $U \sim \text{Exp}(1)$; consider the transformation $W = \log(U)$
- The resulting distribution is known as the standard *extreme value distribution*; we first encountered a more general form of the extreme value distribution in Assignment 4
- For reference, the hazard and density of the standard extreme value distribution are

$$\lambda(w) = e^w$$
$$f(w) = \exp(w - e^w);$$

note that the extreme value distribution is not restricted to be positive, unlike the other distributions we have considered in this course

Weibull and extreme value, part I

- Similarly, suppose $U \sim \text{Exp}(1)$, $T = U^\sigma$, and $Y = \log T$
- In this case, $Y = \sigma W$, where W follows a standard extreme value distribution
- In other words, on the log scale the Weibull distribution with rate $\lambda = 1$ corresponds to a scale family for the extreme value distribution

Exponential and extreme value revisited

- Now let's consider the more general case where $X \sim \text{Exp}(\tau)$; what distribution does $Y = \log(X)$ have?
- We have

$$\lambda(y) = \tau e^y,$$

or, letting $\mu = -\log \tau$,

$$\lambda(y) = e^{y-\mu}$$

- Recall $\lambda(w) = e^w$ for the extreme value distribution; thus, $Y = \mu + W$
- In other words, on the log scale the exponential distribution represents a location family for the EV distribution

Weibull and extreme value, part II

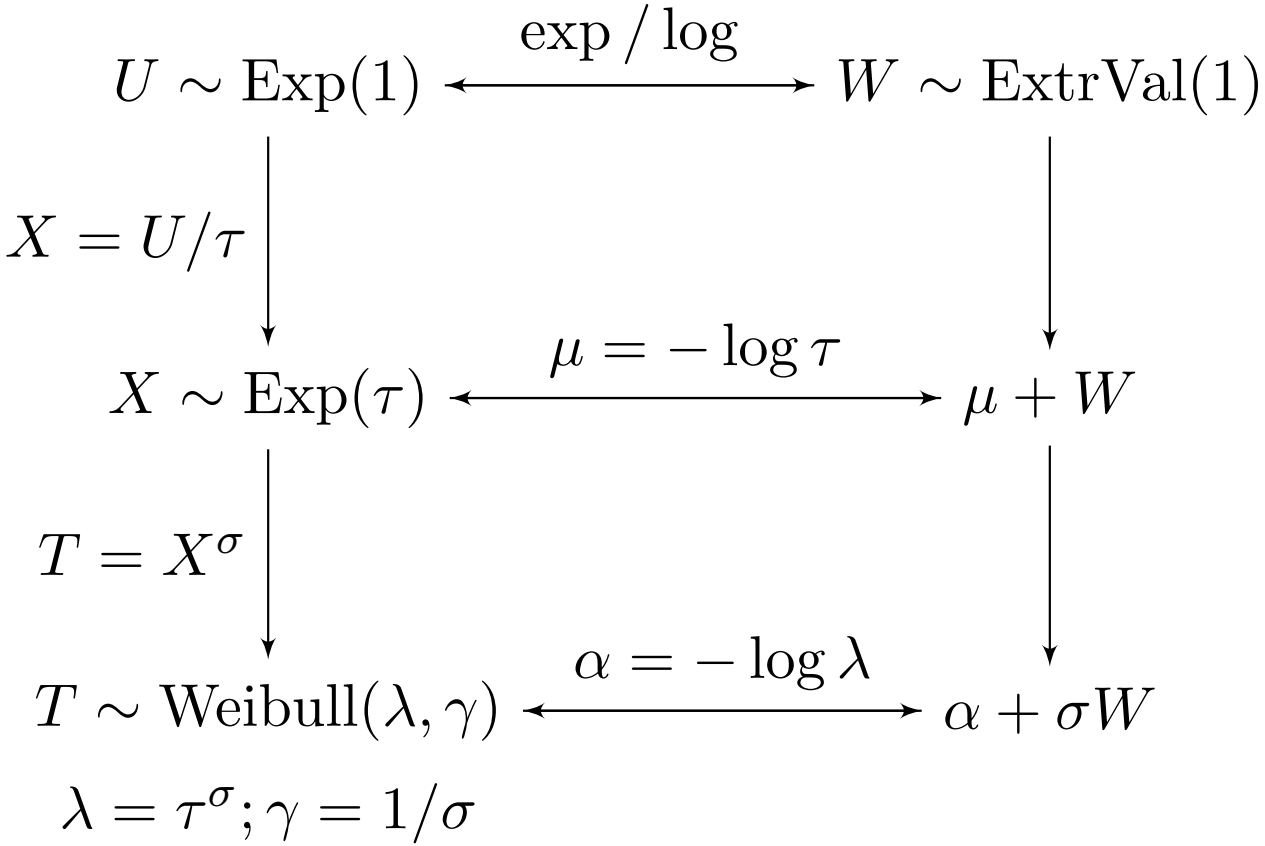
- Finally, for the general case in which $T \sim \text{Weibull}(\lambda, \gamma)$, we have for $Y = \log T$

$$Y = \alpha + \sigma W,$$

  where $\alpha = -\log \lambda$ and $\sigma = 1/\gamma$
- Thus, there is a rather elegant connection between the exponential distribution, the Weibull distribution, and the extreme value distribution
- Furthermore, as we will see shortly, this location-scale relationship will motivate another class of survival regression models

Diagram

Summarizing these relationships:



Weibull regression: Proportional hazards version

- Using the Weibull distribution as the base distribution in a proportional hazards model, we have

$$\lambda_i(t) = \lambda\gamma(t\lambda)^{\gamma-1}\exp(\mathbf{x}_i^T\boldsymbol{\beta})$$

- As with exponential regression, we have an identifiability problem with λ if our model contains an intercept; we therefore have to remove one of them in order to fit the model

Alternate expression of the PH model

- For the sake of illustration, let's keep λ and assume $\boldsymbol{\beta}$ does not contain an intercept
- Another way of thinking about our proportional hazards model is that it assumes $T_i|\mathbf{x}_i \perp\!\!\!\perp \text{Weibull}(\lambda_i, \gamma)$, where

$$\lambda_i = \lambda \exp(\eta_i^*)$$
$$\eta_i^* = \mathbf{x}_i^T \boldsymbol{\beta}/\gamma$$

- In other words, every subject has the same shape parameter γ, but different rate parameters λ_i depending on their covariates

PH model on the log scale

- Thus, in terms of $Y_i = \log T_i$, our model assumes that $Y_i = \alpha_i + \sigma W_i$, where

$$\alpha_i = -\log \lambda - \eta_i^*$$

and the W_i terms follow independent standard extreme value distributions

- In other words, on the log scale our proportional hazards model assumes that

$$Y_i = \alpha + \mathbf{x}_i^T \boldsymbol{\beta}^* + \sigma W_i,$$

where $\alpha = -\log \lambda$, $\boldsymbol{\beta}^* = -\sigma \boldsymbol{\beta}$, $\sigma = 1/\gamma$

Final remarks

- This last expression is obviously rather intriguing, in that it looks exactly like linear regression, only with the error term following a standard extreme value distribution rather than a standard Gaussian distribution
- In fact, one can imagine a general class of survival regression models based on formulas like the one on the previous slide, but with different error distributions
- Regression models of this type are known as accelerated failure time models, and will be the subject of our next lecture, where we will also discuss Weibull regression in further detail