♣ **Gamma** distribution:

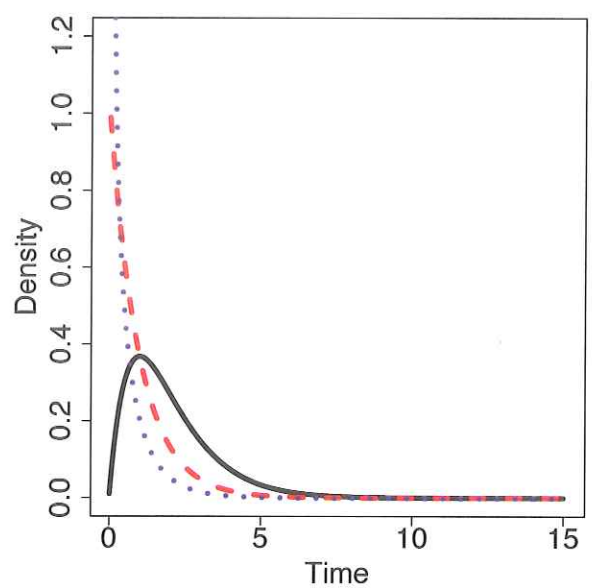
$$y_i \mid \alpha, \gamma \overset{iid}{\sim} \text{Ga}(\alpha, \gamma), \ \alpha > 0 \text{ and } \quad -\infty < \lambda = \log(\gamma) < \infty.$$

- scale parameter $\gamma > 0$ and shape parameter $\alpha > 0$.
- The density function is

$$f(y \mid \alpha, \lambda) = \frac{1}{\Gamma(\alpha)} y^{\alpha-1} \exp(\alpha\lambda - \exp(\lambda) y), \qquad y > 0$$

- $\alpha = 1 \quad \Rightarrow$ exponential distribution
- $\alpha \to \infty \quad \Rightarrow$ a normal distribution.
- $\gamma = 1/2 \quad \Rightarrow$ the chi-square distribution with 2α d.f.
- For α, an integer, we obtain the Erlangian distribution.

- Densities of Gamma distribution



⋆⋆: $(\alpha, \gamma) = (2.0, 1)$ for black solid, (1.0, 1) for red dashed and (0.5, 1) for blue dotted.

$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} \, dx$

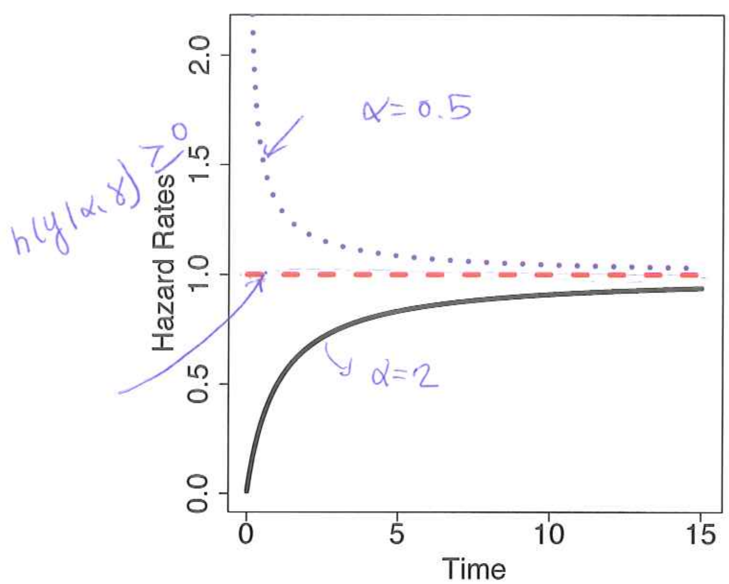
- The survival function $S(y \mid \alpha, \gamma)$

$$
\begin{aligned}
S(y \mid \alpha,\gamma) &= \frac{1}{\Gamma(\alpha)} \int_y^\infty \gamma(\gamma u)^{\alpha-1} \exp(-\gamma u) du \\
= 1 - F(y \mid \alpha, \gamma) \quad & \\
&= 1 - \frac{1}{\Gamma(\alpha)} \int_0^y \gamma(\gamma u)^{\alpha-1} \exp(-\gamma u) du \\
&= 1 - \frac{1}{\Gamma(\alpha)} \int_0^{\gamma y} (v)^{\alpha-1} \exp(-v) dv \\
&= 1 - \mathrm{IG}(\alpha, \gamma y),
\end{aligned}
$$

where IG is the incomplete gamma function.

- The hazard function $h(y \mid \alpha, \gamma) = \dfrac{f(y \mid \alpha, \gamma)}{S(y \mid \alpha, \gamma)}$
  - $\alpha > 1$: monotone increasing with $h(0) = 0$ and $h(y \mid \alpha, \gamma) \to \gamma$ as $y \to \infty$.
  - $\alpha < 1$: monotone decreasing with $h(0) = \infty$ and $h(y \mid \alpha, \gamma) \to \gamma$ as $y \to \infty$

- The hazard rate is $f(y \mid \alpha, \gamma)/S(y \mid \alpha, \gamma)$



⋆⋆: $(\alpha, \gamma) = (\underbrace{2.0, 1})$ for black solid, (1.0, 1) for red dashed and (0.5, 1) for blue dotted.

$\underbrace{\alpha = 1}_{\text{Exp.}}$

- **Gamma** distribution (contd)

$$f(y \mid \alpha, \lambda) = \frac{1}{\Gamma(\alpha)} y^{\alpha-1} \exp(\alpha\lambda - \exp(\lambda)y), \qquad y > 0$$

⋆⋆ No joint conjugate prior for α and λ

⋆⋆ Joint prior for α and λ: assume a priori independence

$$\alpha \sim \text{Gamma}(\alpha_0, \kappa_0), \qquad \text{and} \qquad \lambda = \log(\gamma) \sim \text{N}(\mu_0, \sigma_0^2).$$

- The posterior of α and λ is given by

$$\pi(\alpha, \lambda \mid \tilde{y}, \nu) \propto \prod_{i=1}^{n} \left\{ \frac{1}{\Gamma(\alpha)} \tilde{y}_i^{\alpha-1} \exp(\alpha\lambda - \exp(\lambda)\tilde{y}_i) \right\}^{\nu_i}$$

$$\times \left\{ 1 - IG(\alpha, \exp(\lambda)\tilde{y}_i) \right\}^{1-\nu_i}$$

$$\times \alpha^{\alpha_0 - 1} \exp(-\kappa_0 \alpha) \exp\left(-\frac{(\lambda - \mu_0)^2}{2\sigma_0^2}\right)$$

$$\tilde{y}_i = \begin{cases} y_i & \text{if } \nu_i = 1 \text{ (observed)} \\ \boxed{c_i \quad \text{if } \nu_i = 0 \text{ (censored)}} \end{cases}$$

†† Assume that censoring is *noninformative.* In other words,

- ⋆⋆ potential censoring time is unrelated to the potential event time
- ⋆⋆ Inferences on survival do not depend on the censoring process.

- Can we use this to facilitate the computation?
  - ⋆⋆ Think of this as a type of missing data problem.
  - ⋆⋆ Treat survival time for subjects with being censored as parameters in the model (i.e. data augmentation)
  - ⋆⋆ Sample survival time for those in an MCMC simulation

- Steps in Gibbs sampling become...

We introduce y^* (complete data)

$y_i^* = \tilde{y}_i$ if $\nu_i = 1$ (survival time is observed)

$y_i^* > c_i (= \tilde{y}_i)$ if $\nu_i = 0$ (survival time is censored)

y_i^* → latent parameters; c_i known ⇒ will impute

$y_i^* | \alpha, \lambda \overset{iid}{\sim} Ga(\alpha, \lambda)$

$$f(y_i^* \mid \alpha, \lambda) = \frac{1}{\Gamma(\alpha)} \exp\left((\alpha-1) \log(y_i^*) + \alpha\lambda - \exp(\lambda) y_i^* \right)$$

$$P(\alpha, \lambda, y^*, D) = \prod_{i=1}^{n} \left\{ P(y_i^* \mid \alpha, \lambda) \cdot \underline{\left(I(y_i^* > y_i) \right)}^{1-\nu_i} \right\}$$

$$\times \; \pi(\alpha) \; \pi(\lambda)$$

What do we simulate?

$\alpha, \lambda, \; y_i^*$ w/ $\nu_i = 0$

$\Rightarrow$ full conditionals

① $\pi(\alpha \mid y_i^*, \lambda) \propto \prod_{i=1}^{n} p(y_i^* \mid \alpha, \lambda) \cdot \pi(\alpha)$

② $\pi(\lambda \mid y_i^*, \alpha) \propto \prod_{i=1}^{n} p(y_i^* \mid \alpha, \lambda) \cdot \pi(\lambda)$

③ $\pi(y_i^* \mid \alpha, \lambda) \propto p(y_i^* \mid \alpha, \lambda) \cdot I(y_i^* > y_i)$

• $\boxed{i \text{ w/ } \nu_i = 0}$

♣ We have:

Original Scale $t > 0$	Log Scale $-\infty < y < \infty$
Weibull	Extreme Value
LN	N
Log-Logistic	Logistic

$y = \log(t)$

♣ **Log-logistic** distribution:

$$T_i \mid \mu, \sigma \overset{iid}{\sim} \mathrm{Log-Logistic}(\mu, \sigma), \quad T_i > 0.$$

That is,

$$f(t \mid \alpha, \lambda) = \frac{\lambda \alpha t^{\alpha - 1}}{(1 + t^{\alpha} \lambda)^2},$$

density for log - logistic

Here $0 < \alpha$ and $0 < \lambda$.

• Connection to the logistic distribution: $Y = \log(T)$ follows the logistic distribution. $\sigma = \frac{1}{\alpha}$ $\log(\lambda) = -\frac{\mu}{\sigma}$ $\mu = -\sigma \log(\lambda)$

$= -\frac{1}{\alpha} \log(\lambda)$

• Let $\alpha = 1/\sigma$ and $\lambda = \exp(-\mu/\sigma)$.

• The density function of $y = \log(t)$ (logistic distribution) is

$$f(y \mid \mu, \sigma) = \frac{\exp(\frac{y-\mu}{\sigma})}{\sigma(1 + \exp(\frac{y-\mu}{\sigma}))^2}, \quad -\infty < y < \infty, \Bigg) \text{ density for logistic}$$

where μ and σ are the mean and scale parameter of Y.

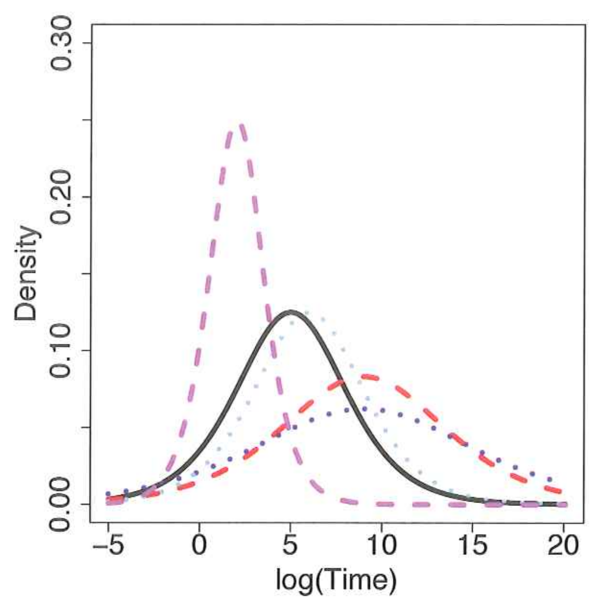
• The logistic distribution closely resembles the normal distribution but the survival function is mathematically more tractable.

- The survival function of t (log-logistic) is

$$S(t \mid \alpha, \lambda) = \frac{1}{1 + \lambda t^{\alpha}},$$

where $\alpha = 1/\sigma > 0$ and $\lambda = \exp(-\mu/\sigma)$.

• The logistic distribution closely resembles the normal distribution but the survival function is mathematically more tractable.

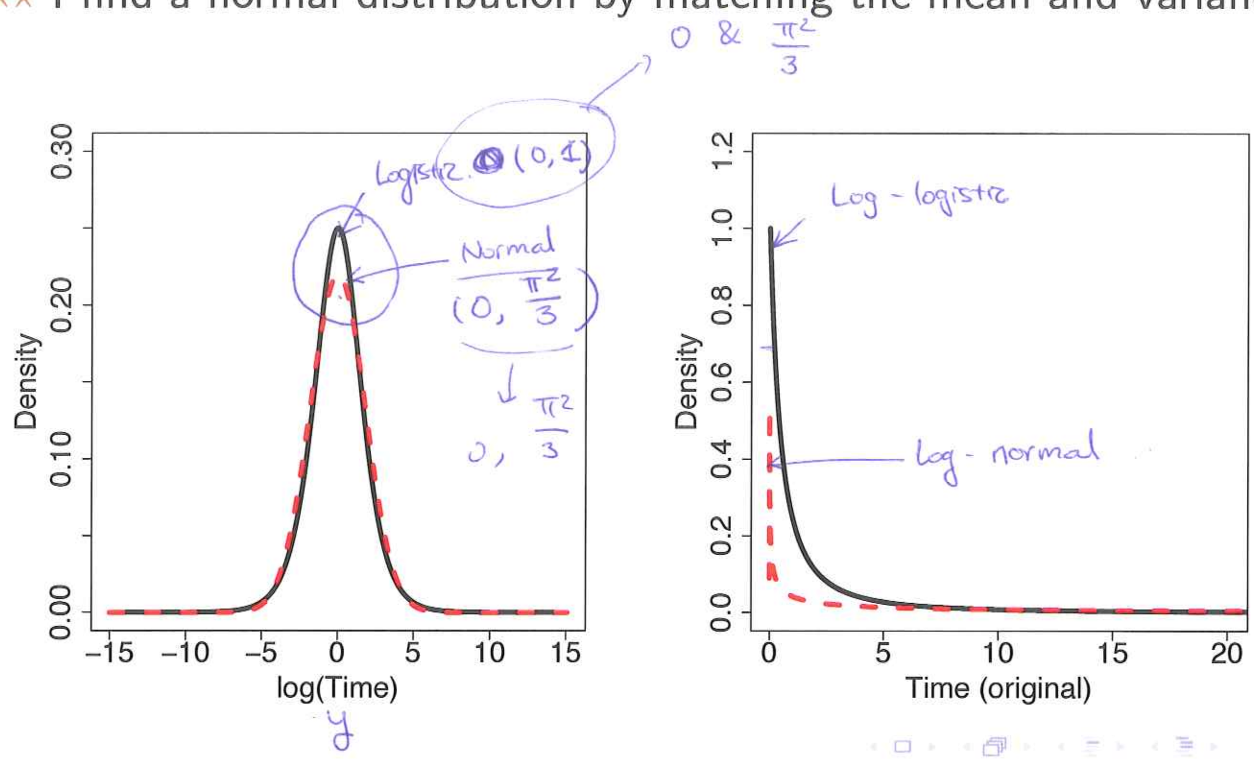


⋆⋆: $(\mu, \sigma) = (5, 2)$, (9, 3), (9, 4), (6, 2), (2, 1) for black, red, blue, cyan, magenta.

• The logistic distribution and normal distribution may imply something different for survival time on the original scale.

⋆⋆ $(\mu, \sigma) = (0, 1)$ for the logistic distribution $\Rightarrow$ mean $= \mu$ and var$=\boxed{\sigma^2\pi^2/3}$.

⋆⋆ I find a normal distribution by matching the mean and variance.



- The hazard function is

$$h(t \mid \alpha, \lambda) = \frac{\alpha \lambda t^{\alpha-1}}{1 + \lambda t^{\alpha}}, \quad = \frac{f(t)}{S(t)}$$

where $\alpha = 1/\sigma > 0$ and $\lambda = \exp(-\mu/\sigma)$.

  - $\alpha \leq 1$: monotone decreasing
  - $\alpha > 1$: the hazard rate increases initially to a maximum at time $((\alpha - 1)/\lambda)^{1/\alpha}$ and then decreases to zero as time approaches infinity.

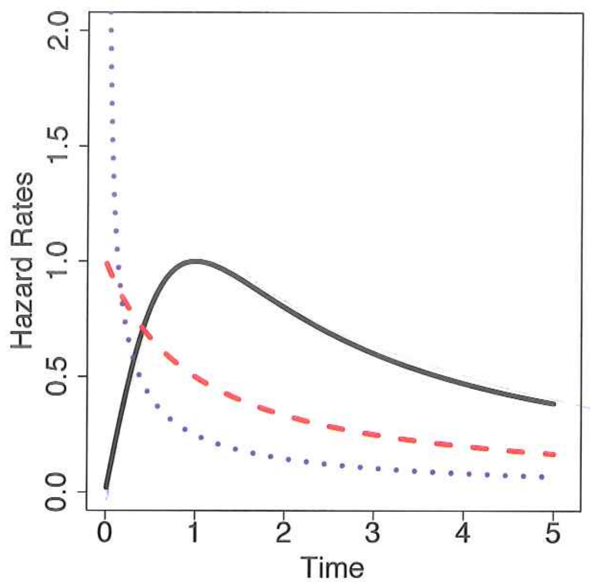
hump-shaped

- Similar to the Weibull and exponential models, it has a simple expression for $h(t \mid \alpha, \lambda)$ and $S(t \mid \alpha, \lambda)$.

- Its hazard rate is similar to the log normal except in the extreme tail of the distribution, but its advantage is its simpler hazard function and survival function.

- The hazard rate is

$$h(t \mid \alpha, \lambda) = \frac{\alpha \lambda t^{\alpha-1}}{1 + \lambda t^{\alpha}},$$



⋆⋆: $(\alpha, \lambda) = (2.0, 1)$ for black solid, (1.0, 1) for red dashed and (0.5, 1) for blue dotted.

- The density function of $y = \log(t)$ (logistic distribution) is

$$f(y \mid \mu, \sigma) = \frac{\exp(\frac{y-\mu}{\sigma})}{\sigma(1+\exp(\frac{y-\mu}{\sigma}))^2}, \quad -\infty < y < \infty,$$

where μ and σ are the mean and the scale parameter of Y.

- Writing the model in a general linear model format, $y = \mu + \sigma w$.

$\Rightarrow \quad w = \frac{y-\mu}{\sigma} \qquad dw = \frac{1}{\sigma} dy$

$f(w) = \frac{e^w}{\cancel{\sigma}(1+e^w)^2} \cdot \cancel{\sigma} = \frac{e^w}{(1+e^w)^2},$

$\Rightarrow$ standard logistic distribution

replace μ with $x\beta$.

AMS 276
Lec 3: Accelerated Failure-Time Model

Fall 2016

$y_i | \alpha, \lambda \sim W(\alpha, \lambda)$ $\qquad$ $y_i \mid \alpha, x_i \overset{indep}{\sim} W(\alpha, \lambda_i)$

a function of x_i and β

- When comparing two or more groups of time-to-event, we may estimate the survival function for each group.
- Often additional information on subjects that may affect their outcome are collected.

  e.g.

  ⋆⋆ demographic variables: age, sex, socioeconomic status or education

  ⋆⋆ physiological variables: blood pressure, blood glucose levels
- Let $\mathbf{X}_i$ denote a $p-$dim vector of covariates for subject i.
- Covariates may be constant (or fixed) values known at time 0, such as initial disease status.
- They may be time dependent, i.e., their value changes over time such as current disease status, serial blood pressure measurements (will be discussed later).

♣ Regression Models for Survival Data

⋆⋆ Often interested in studying the relationship between the failure time (T) and covariates ($\mathbf{X}$: $p \times 1$ associated with T).

e.g. Predict the distribution of the failure time from a set of covaraites.

⋆⋆ Adjust the survival function to account for covariates.

- Two Common Approaches:

⋆⋆ **Accelerated Failure-Time Model**

⋆⋆ Proportional Hazards Model (Multiplicative Hazards Model - Cox-type model). Will be discussed next.

♣ Approach 1: Accelerated Failure-Time Model (KM, chapter 12 & ICS 10.2)

ICS Chapter 2

- Suppose we have right censored survival data.

- We consider the linear relationship between log time and co-variates.

  $\Rightarrow$ Let $Y = \log(T) \in (-\infty, \infty)$ (bad notation!) and assume a linear model for Y;

$$Y = -\beta_0 - \boldsymbol{\beta}'\mathbf{X} + \sigma W.$$

  ⋆⋆ $\mathbf{X} = (X_1, \ldots, \mathbf{X}_p)$: a vector of known & fixed time explanatory covarites

  ⋆⋆ $(\beta_0, \boldsymbol{\beta})$: a vector of $p + 1$ unknown regression coefficients.

  ⋆⋆ W: the error term

  $\Rightarrow$ Will place a distribution on W

  $\Rightarrow$ It will determine the distr. of Y

  ($\Rightarrow$ the distr. of T)

- Why is it called the accelerated failure-time model?

⋆ Recall survival time $T = \exp(Y)$ and $Y = -\beta_0 - \beta'\mathbf{X} + \sigma W$.

⋆ First, if $\mathbf{X} = \mathbf{0}$, then $Y = -\beta_0 - \beta'\mathbf{X} + \sigma W = -\beta_0 + \sigma W$

⋆ Now, look at the survival function for a subject with $\mathbf{X}$.

$$
\begin{aligned}
S(t \mid \mathbf{X}) &= \Pr(T > t \mid \mathbf{X}) \\
&= \Pr(Y > \log(t) \mid \mathbf{X}) \\
&= \Pr(-\beta_0 - \beta'\mathbf{X} + \sigma W > \log(t)) \\
&= \Pr(-\beta_0 + \sigma W > \log(t) + \beta'\mathbf{X}) \\
&= \Pr(\exp(-\beta_0 + \sigma W) > t\exp(\beta'\mathbf{X})) \\
&= S_0(t\exp(\beta'\mathbf{X})),
\end{aligned}
$$

The prob. of surviving beyond time t for a ~~sjt~~ subject w/ X.

Y w/ $\cancel{X}$=0

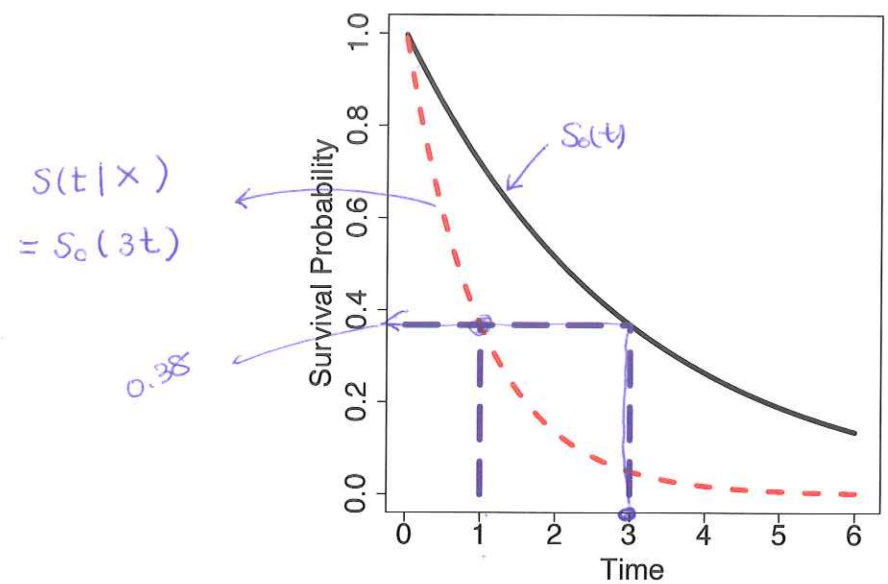
where $S_0(t)$ is the survival function of $T = \exp(Y)$ for $\mathbf{X} = \mathbf{0}$.

The prob. of surviving beyond time $t \cdot e^{\beta X}$ for a subject w/ X = 0

- We have $S(t \mid \mathbf{X}) = S_0(t \exp(\beta'\mathbf{X}))$.

$\star$ Suppose that $\underbrace{\exp(\beta'\mathbf{X})}_{=\text{ acceleration factor}} = 3 \Rightarrow S(t \mid \mathbf{X}) = S_0(3 \cdot t)$.

$\leftrightarrow$ the probability that a subject having $\mathbf{X}$ survives longer than time t is the same as that of a subject having $\mathbf{X} = \mathbf{0}$ longer than time $(3 \cdot t)$.



- We have

$$S(t \mid X) = \Pr(T > t \mid \mathbf{X}) = S_0(t \exp(\beta' \mathbf{X})),$$

where $S_0(t)$ is the survival function of $T = \exp(Y)$ for $\mathbf{X} = \mathbf{0}$

⋆⋆ Observe that the survival function of an individual with covariate $\mathbf{X}$ at time t is the same as the survival function of an individual with a baseline survival function at a time $t \exp(\beta' \mathbf{X})$.

$X=0$ $S_0(t)$

(X)

⋆⋆ That is, the effect of the explanatory variables in the original time scale is to change the time scale by a factor $\exp(\beta' \mathbf{X})$ (acceleration factor).

$e^{\beta X} > 1$

⋆⋆ The time is either accelerated by a constant factor or degraded by a constant factor.

$e^{\beta X} < 1$

- $S(t \mid X) = S_0(t e^{X\beta})$
- $f(t \mid X) = -\dfrac{d S(t \mid X)}{dt}$

- How about other quantities?
- Observe

$$f(t \mid \mathbf{X}) = f_0\{t \exp(\beta'\mathbf{X})\} \exp(\beta'\mathbf{X}).$$

- Express the conditional hazard rate of an individual with $\mathbf{X}$ as

$$h(t \mid \mathbf{X}) = h_0\{t \exp(\beta'\mathbf{X})\} \exp(\beta'\mathbf{X}).$$

$$= \frac{f(t \mid X)}{S(t \mid X)} = \frac{f_0(t e^{X\beta})\, e^{X\beta}}{S_0(t \cdot e^{X\beta})}$$

$\rightarrow h_0(t \cdot e^{X\beta})$

⋆⋆ This shows the relationship of the hazard rate for an individual with covariate $\mathbf{X}$ with a baseline hazard rate.

- How about other quantities? (contd)
- Observe

$$f(t \mid \mathbf{X}) = f_0\{t\exp(\beta'\mathbf{X})\}\exp(\beta'\mathbf{X}).$$

- $\mathrm{E}(T \mid \mathbf{X}) = \exp(\beta'\mathbf{X})\mathrm{E}(T \mid \mathbf{X} = \mathbf{0})$

★★ Consider two individuals, i_1 and i_2 with identical covariate vectors $\mathbf{X}_{i_1}$ and $\mathbf{X}_{i_2}$ except one covariate $X_{ki_2} = X_{ki_1} + 1$.

★★ Observe

1

$$\frac{\mathrm{E}(T \mid \mathbf{X}_{i_1})}{\mathrm{E}(T \mid \mathbf{X}_{i_2})} = \frac{\exp(\beta'\mathbf{X}_{i_1})\mathrm{E}(T \cancel{\mid \mathbf{X} = \mathbf{0}})}{\exp(\beta'\mathbf{X}_{i_2})\mathrm{E}(T \cancel{\mid \mathbf{X} = \mathbf{0}})} = \exp(-\beta_k)$$

2

★★ $\exp(-\beta_k)$: multiplicative change in mean survival time associated with 1 unit increase in X_k, while keeping all other X's fixed.

- How about other quantities? (contd)
- Observe

$$S(t \mid \mathbf{X}) = S_0\{t \exp(\beta'\mathbf{X})\}.$$

- Percentiles:

- t_p, t_{0p} : p-th percentile for the distribution of T w/ X and T w/ $X=0$

$F(t_p \mid x) = \frac{p}{100}$

- $\left.\begin{aligned} S(t_p \mid x) &= 1 - \frac{p}{100} \\ S_0(t_{0p}) &= 1 - \frac{p}{100} \end{aligned}\right\} \Rightarrow S(t_p \mid x) = S_0(t_{0p})$

$\Rightarrow S(t_p \mid x) \underset{\text{AFT}}{=} S_0(t_p \cdot e^{x\beta}) = S_0(t_{0p})$

$\Rightarrow t_p \cdot e^{x\beta} = t_{0p}$

♣ What is coming?

- Let $Y = \log(T) \in (-\infty, \infty)$ and assume a linear model for Y;

$$Y = -\beta_0 - \boldsymbol{\beta}'\mathbf{X} + \sigma W.$$

- What would we do with the error term (W)?
  - Assume the distribution of the error term is a member of some parametric family.
    - ⋆⋆ $W \sim \mathrm{V} \Rightarrow T \sim \mathrm{Weibull}$
    - ⋆⋆ $W \sim \text{Standard Logistic} \Rightarrow T \sim \mathrm{Log-Logistic}$
    - ⋆⋆ $W \sim \mathrm{N}(0,1) \Rightarrow T \sim \mathrm{LN}$
  - Assume the error terms are iid from some unknown distribution F. $\Rightarrow$ Nonparametric!

$W \sim F \quad F \sim DP$

♣ Parametric Approach 1: Std Extreme Value Distribution for W

○ Recall!

- Survival time $T \Rightarrow Y = \log(T)$
- Let $Y \mid \alpha, \lambda \sim V(\alpha, \lambda)$.

$\lambda = -\mu\alpha = -\frac{\mu}{\sigma}$ $\quad \alpha = \frac{1}{\sigma}$

- The distribution of $W = \frac{Y-\mu}{\sigma}$ where $\mu = -\frac{\lambda}{\alpha}$ and $\sigma = \frac{1}{\alpha}$?

$\Rightarrow$ W follows the standard extreme value distribution.

○ Let's consider our $\underbrace{Y = -\beta_0 - \boldsymbol{\beta}'\mathbf{X} + \sigma W}$ where $W \sim$ std V.

$\Rightarrow$ What is the distribution of our Y?

$$W = \frac{Y + \beta_0 + \beta' x}{\sigma} \qquad \mu = -\beta_0 - \beta' x$$

$$Y \mid \alpha, \lambda \sim V(\alpha, \lambda)$$

$$\alpha = \frac{1}{\sigma}, \quad \lambda = -\frac{-\beta_0 - \beta' x}{\sigma}$$

$$= \alpha \cdot (\beta_0 + \beta' x)$$

$$Y_i \mid \alpha, \lambda_i \sim V\left(\alpha = \frac{1}{\sigma}, \quad \lambda_i = \frac{1}{\sigma} \cdot (\beta_0 + \beta' x_i)\right)$$

- Recall! Let $Y \mid \alpha, \lambda \sim \mathrm{V}(\alpha, \lambda)$.

  ⋆⋆ pdf:

$$f(y \mid \alpha, \lambda) = \alpha \exp(\lambda + \alpha y - \exp(\lambda + \alpha y)), \quad -\infty < y < \infty.$$

  ⋆⋆ Survival function: $S_Y(y \mid \alpha, \lambda) = \exp(-\exp(\lambda + y\alpha))$

- Now consider our $Y = -\beta_0 - \boldsymbol{\beta}'\mathbf{X} + \sigma W$.
- How to write down the likelihood for right-censored data in Y?

$\alpha = \frac{1}{\sigma}, \qquad \lambda_i = \frac{1}{\sigma}(\beta_0 + \beta' x_i)$

$$\mathcal{L}(\beta_0, \beta; \sigma \mid D) = \prod_{i=1}^{n} \left\{ f(y_i \mid \beta_0, \beta, \sigma, x_i) \right\}^{\nu_i} \left\{ S(y_i \mid \beta_0, \beta, \sigma, x_i) \right\}^{1-\nu_i}$$

$$= \prod_{i=1}^{n} \left\{ \frac{1}{\sigma} \exp\left(\frac{\beta_0 + \beta' x_i}{\sigma} + \frac{y_i}{\sigma} - \exp\left(\frac{\beta_0 + \beta' x_i}{\sigma} + \frac{y_i}{\sigma} \right) \right) \right\}^{\nu_i}$$

$$\times \left\{ \exp\left(- \exp\left(\frac{y_i + \beta_0 + \beta_1 x_i}{\sigma} \right) \right) \right\}^{1-\nu_i}$$