Survival Analysis
Math 434 – Fall 2011

Part III: Chap. 2.5,2.6 & 12

Jimin Ding
Math Dept.
www.math.wustl.edu/ jmding/math434/index.html

Parametric Models
Outlines ................................................................. slide #3
Exponential Distribution ......................................... slide #4
Weibull Distribution ................................................ slide #5
Lognormal Distribution .......................................... slide #6
Gamma Distribution ............................................... slide #7
Log-logistic Distribution ....................................... slide #8
Gompertz Distribution ........................................... slide #9
Parametric Regression Models with Covariates ............ slide #10
Accelerated Failure-Time (AFT) Model ...................... slide #11
Proportional Hazards Model .................................... slide #12
Proportional Odds Model ....................................... slide #13
Model Comparison Using Akaikie Information Criterion (AIC) .... slide #14
Outlines

- Exponential distribution \( \text{exp}(\lambda), \lambda > 0; \)
- Weibull distribution \( \text{Weibull}(\lambda, \alpha), \lambda > 0; \)
- Lognormal distribution \( \text{logN}(\mu, \sigma^2), \sigma > 0; \)
- Gamma distribution \( \text{Gamma}(\lambda, \beta), \lambda > 0; \)
- Log-logistic distribution \( \text{loglogit}(\lambda, \alpha), \lambda > 0; \)
- * Gompertz distribution \( \text{Gompertz}(\theta, \alpha), \alpha > 0. \)

For each parametric model, we will discuss the distribution properties and parameter estimation.
Parameter Estimation in Exponential Model
Weibull Distribution

Extension of exponential distribution:

![Graphs showing survival, density, and hazard functions with different values of lambda and alpha.](image-url)
Weibull Distribution Properties
Parameter Estimation in Weibull Model
Lognormal Distribution

Exponential of normal distribution:

- Survival Function
- Density Function
- Hazard Function

Jimin Ding, October 4, 2011
Survival Analysis, Fall 2011 — slide #6
Parameter Estimation in Lognormal Model
Gamma Distribution

Survival Function

Density Function

Hazard Function

Jimin Ding, October 4, 2011
Survival Analysis, Fall 2011 — slide #7
Gamma Distribution Properties
Parameter Estimation in Gamma Model
Log-logistic Distribution

Survival Function

Density Function

Hazard Function
Log-logistic Distribution Properties
Parameter Estimation in Logistic Model
Parametric Regression Models with Covariates

Denote $X$ as the survival time of interest, $Z = (Z_1, \ldots, Z_p)$ as the vector of covariates. Consider a linear model for modeling $Y = \log(X)$, namely,

$$Y = \mu + \gamma^T Z + \sigma W,$$

where $\gamma^T = (\gamma_1, \ldots, \gamma_p)$ is a vector of regression coefficients and $W$ is the error distribution. Common choices for $W$:

- $W \sim N(0, 1) \Rightarrow X \sim \log N(\mu + \gamma^T Z, \sigma^2)$;
- $W \sim$ extreme value distribution $\Rightarrow X \sim$ Weibull.
- $W \sim$ standard logistic $\Rightarrow X \sim$ log-logistic.
Accelerated Failure-Time (AFT) Model

The above regression models are special cases of accelerated failure-time model (AFT). AFT model states that the survival function of an individual with covariate $Z$ could be written as the survival function of $\exp(\mu + \sigma W)$ with scaled age $x \exp(-\gamma^T Z)$. Mathematically we have

$$S(x|Z) = S_0(x \exp(-\gamma^T Z)),$$

where $S_0(x)$ is the survival function of the random variable $\exp(\mu + \sigma W)$ and referred as the baseline survival function. When $\gamma$ is negative, the time is accelerated by a constant.

The AFT property is independent of error distribution. A general AFT model relax the error distribution assumption and belongs semiparametric models. (The $(\gamma)$ is a p-dim parameter and the distribution of $W$ is an infinite dimensional parameter.)

Proportional Hazards Model

Another common model for survival time assumes

$$h(x|Z) = h_0(x) \exp(\beta^T Z),$$

where $h_0(x)$ is called the baseline hazard function. This type of model has constant hazard ratios over time.

If the form of $h_0(x)$ is known, then it is corresponding to a parametric model. For example $h_0(x) = \alpha \lambda t^{\alpha-1}$, the proportional hazards model is same as Weibull model. Note that $\beta = -\gamma/\sigma$. In fact, Weibull distribution is the only distribution satisfies both proportional Hazards assumption and AFT assumption.

If the form of $h_0(x)$ is unknown and need to be estimated, then it is a semiparametric model. This model was first proposed by Cox (1972) and is one of the most popular survival models currently.
Proportional Odds Model

Define \( \text{odds}(t) = \frac{s(t)}{1-s(t)} \). The proportional odds model assumes:

\[
\text{odds}(t|Z) = \text{odds}(t|Z = 0) \exp(-\beta^T Z),
\]

and \( \text{odds}(t|Z = 0) \) is called the baseline odds function. This type of model has constant odds ratios over time.

Similar as the proportional hazard model, if the form of \( \text{odds}_0(t) = \text{odds}(t|Z = 0) \) is unknown, this is a semiparametric model.

Log-logistic distribution is the only distribution satisfies both proportional odds assumption and AFT assumption. The parameters of log-logistic distribution are

\[
\lambda = \exp(\mu/\sigma), \quad \alpha = 1/\sigma, \quad \beta_i = -\gamma_i/\sigma.
\]

Model Comparison Using Akaike Information Criterion (AIC)

To compare nested parametric models, we could use Wald or Likelihood ratio tests.

In general, to compare two parametric models, one may use AIC defined as,

\[
AIC = -2 \log (\text{Likelihood}) + 2p,
\]

where \( p \) is the total number of parameters in the model. For example, \( p = 1 \) for the exponential model, \( p = 2 \) for the Weibull model and \( p = 3 \) for the generalized gamma model.

The smaller AIC, the better the model is. Example 12.1 on P407.