Math 659: Survival Analysis

Chapter 2 — Basic Quantiles and Models (II)

Wenge Guo

July 22, 2011

Review of Last lecture (1)

- A **lifetime** or **survival time** is the time until some specified **event** occurs. This event may be death, the appearance of a tumor, the development of some disease, recurrence of a disease, equipment breakdown, cessation of breast feeding, and so on.
- **Survival random variable:** A random variable X is a survival random variable if an observed outcome x of X is always positive.
- **Several functions** characterize the distribution of a survival random variable: probability density function (pdf) $f(x)$, cumulative distribution function (cdf) $F(x)$ survival function (sf) $S(x)$, hazard function (hf) $h(x)$, cumulative hazard function $H(x)$, and mean residual lifetime $mrl(x)$ at time x, respectively.

Review of Last lecture (2)

Implication of these functions:

- The survival function $S(x)$ is the probability of an individual surviving to time x.
- The hazard function $h(x)$, sometimes termed risk function, is the chance an individual of time x experiences the event in the next instant in time when he has not experienced the event at x.
- A related quantity to the hazard function is the cumulative hazard function $H(x)$, which describes the overall risk rate from the onset to time x.
- The mean residual lifetime at age x, $mrl(x)$, is the mean time to the event of interest, given the event has not occurred at x.

Relationship Summary

We only need to know one of these functions, the rest can be derived. In particular,

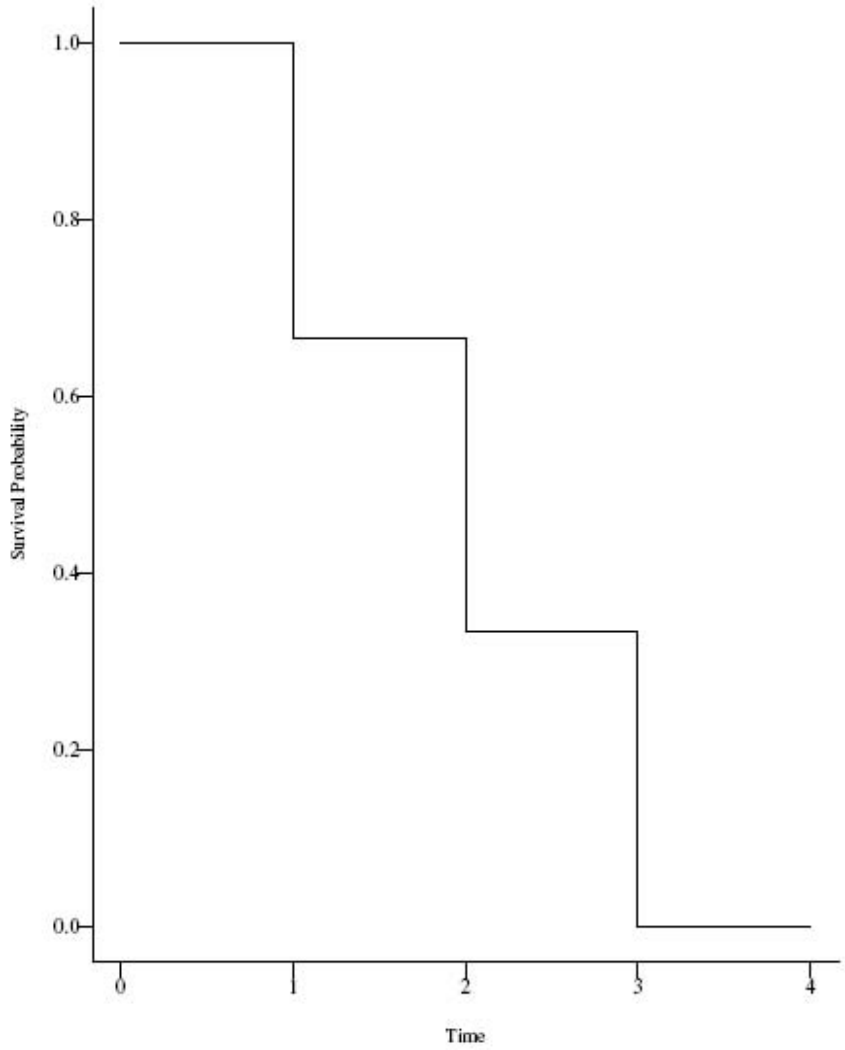
- $S(x) = 1 - F(x), F(x) = \int_0^x f(t)dt, S(x) = \int_x^\infty f(t)dt$
- $f(x) = \frac{dF(x)}{dx} = -\frac{dS(x)}{dx}$
- $h(x) = \frac{f(x)}{S(x)}$, $H(x) = \int_0^x h(t)dt, h(x) = \frac{dH(x)}{dx}$
- $mrl(x) = \frac{\int_x^\infty (t-x)f(t)dt}{S(x)} = \frac{\int_x^\infty S(t)dt}{S(x)}$ and $S(x) = \frac{mrl(0)}{mrl(x)} e^{-\int_0^x \frac{dt}{mrl(t)}}$
- The key relation is $S(x) = exp(-\int_0^x h(t)dt) = exp(-H(x))$.

Discrete Case

- Suppose that X takes values $x_j, j = 1, 2, 3, \cdots$
- The probability mass function (pmf) $p(x_j) = \Pr(X = x_j)$, $j = 1, 2, \cdots$ where $x_1 < x_2 < \cdots < x_j < \cdots$
- The survival function is defined as $S(x) = \Pr(X > x) = \sum_{x_j > x} p(x_j)$
- Example: X with pmf $p(x_j) = \Pr(X = j) = 1/3, j = 1, 2, 3$, then the survival function is

$$S(x) = \Pr(X > x) = \sum_{x_j > x} p(x_j) = \begin{cases} 1, & 0 \le x < 1 \\ 2/3, & 1 \le x < 2 \\ 1/3, & 2 \le x < 3 \\ 0, & x \ge 3 \end{cases}$$

Plot of the Survival Function for Discrete Case



Hazard Function for the Discrete Case

- The hf is defined as
  $h(x_j) = \Pr(X = x_j | X \geq x_j) = \frac{p(x_j)}{S(x_{j-1})}, j = 1, 2, \cdots$ with
  $S(x_0) = 1$
- Because $p(x_j) = S(x_{j-1}) - S(x_j)$, thus
  $h(x_j) = \frac{S(x_{j-1}) - S(x_j)}{S(x_{j-1})} = 1 - \frac{S(x_j)}{S(x_{j-1})}$
- The survival function can be written as the product of the conditional survival probabilities.
  $S(x) = \prod_{x_j \leq x} S(x_j)/S(x_{j-1})$

Relationship between $h(x_j)$ and $S(x)$

- Note that,

$$\begin{aligned} S(x) =& \Pr(X > x) = \Pr(X > x | X > x_j) \Pr(X > x_j) \\ =& \Pr(X > x | X > x_j) \Pr(X > x_j | X > x_{j-1}) \Pr(X > x_{j-1}) \\ =& \cdots \end{aligned}$$

- Because $\Pr(X > x_j | X > x_{j-1}) = S(x_j)/S(x_{j-1})$
- Thus $S(x) = \prod_{x_j \leq x} [1 - h(x_j)]$, which provides the basis for the Kaplan-Meier estimator of the survival function
- The cumulative hazard function is defined as $H(x) = \sum_{x_j \leq x} h(x_j)$
- Note that $S(x) = \exp[-H(x)]$, which could be used to provide an alternative estimator for the survival function

Relationship Summary for Discrete Lifetimes

Interrelationships between the various quantities for discrete lifetimes X may be summarized as

- $S(x) = \sum_{x_j > x} p(x_j) = \prod_{x_j \leq x}[1 - h(x_j)]$,
- $p(x_j) = S(x_j) - S(x_{j-1}) = h(x_j)S(x_{j-1}), j = 1, 2, \ldots,$
- $h(x) = \frac{p(x_j)}{S(x_{j-1})}$,
- $mrl(x) = \frac{(x_{i+1}-x)S(x_i)+\sum_{j \geq i+1}(x_{j+1}-x_j)S(x_j)}{S(x)}$, for $x_i \leq x < x_{i+1}$.

Parametric Distributions

- We mainly use nonparametric and semiparametric models and methods in this class
- Parametric models are useful and widely used in some area, such as reliability data analysis, engineer statistics
- The main purpose of survival analysis is to interpret data, for example, what is the population life distribution, what is treatment effect on the survival distribution, based on data collected
- In reliability, often it is needed to make prediction for the fraction failing of a product after three year based on one-year data. Thus extrapolation is needed and parametric models are used
- In reliability, the use of parametric models can be justified by physical/chemical principles or engineering knowledge, but it is hard to make this justification for human body

The Exponential

- The sf is $S(x) = \exp(-\lambda x), \lambda > 0, x \geq 0$
- The pdf is $f(x) = \lambda \exp(-\lambda x)$
- The hf is λ
- The mean and sd is $1/\lambda$
- The lack of memory property, which means $\Pr(X \geq x + z | x \geq x) = \Pr(X \geq z)$

The Exponential (2)

Property: An important feature of the exponential distribution is the '**memoryless property**', $P(X > x + z|X > x) = P(X > z)$. That is, on reaching any age, the probability of surviving z more units of time is the same as it was at age zero.

Example: For a component with an exponential distributed lifetime, the probability that a one-year-old component lasts 3 more months in operation is the same as the probability that a ten-year-old component lasts 3 more months in operations.

Implication: It indicates that the component's lifetime does not pass through a period of 'old ages', where there is an increased risk of mortality.

The Exponential (3)

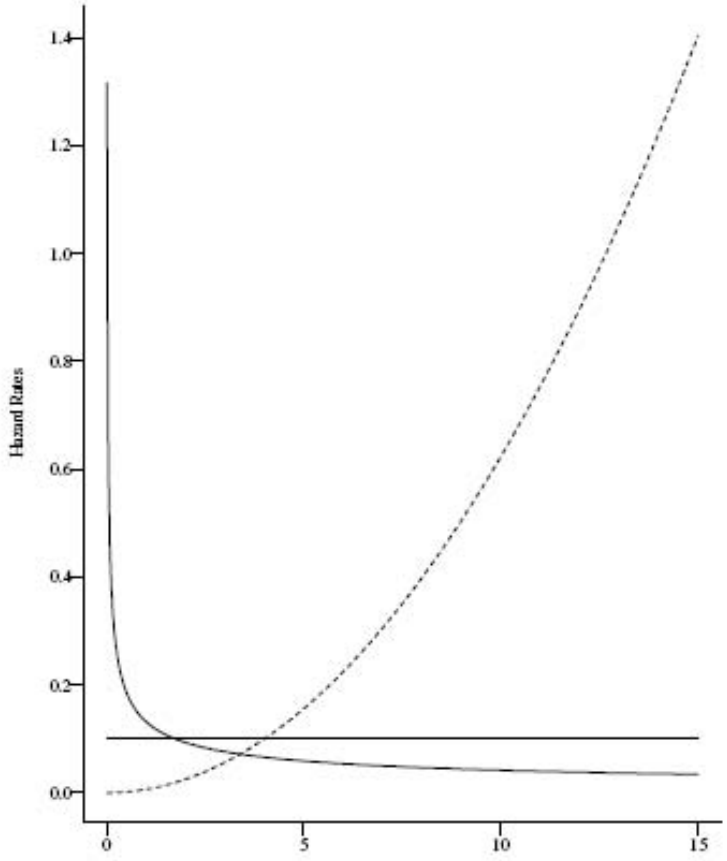
Application: The exponential probability model is one of the most commonly used probability models for modeling lifetimes of components. However, its constant hazard rate appears too restrictive in health and some industrial applications.

Parameter: The parameter β is an important scale parameter, we have the following conclusion: If $X \sim exp(\lambda)$, then $\lambda X \sim exp(1)$.

The Weibull Distribution

- Widely used in reliability
- The sf is $S(x) = \exp(-\lambda x^{\alpha})$
- λ is a scale parameter and α is a shape parameter
- The hf is $\lambda\alpha x^{\alpha-1}$, which is more flexible, allows different shapes of the hazard
- In particular, α determines the shape of the hazard function
  - $\alpha > 1$, increasing
  - $\alpha = 1$, constant
  - $\alpha < 1$, decreasing
- Let $Z = \ln(\lambda X^{\alpha})$, then sf of Z is $\exp[-\exp(z)]$, which is called the standard smallest extreme value distribution

The Weibull Hazard Function



Several Comments on Weibull Model

- The Weibull model has a very simple hazard function and survival function.
- It is a very useful model in many engineering context. Its two parameters make the Weibull a very flexible model in a wide variety of situations: increasing hazards, decreasing hazards, and constant hazards.

Several Comments on Weibull Model (2)

- An understanding of the hazard rate may provide insight as to what is causing the failures:
  - A decreasing hazard rate would suggest "infant mortality". That is, defective items fail early and the failure rate decreases over time as they fall out of the population.
  - A constant hazard rate suggests that items are failing from random events.
  - An increasing hazard rate suggests "wear out" - parts are more likely to fail as time goes on.
- However, the mean and variance of the distribution are more difficult to determine.

The Log Normal Distribution

Since many lifetimes are measured on the logarithm scale, and such transformations often increase the symmetry in data, it is important to examine the **log normal distribution** where the transformed data are normally distributed.

A positive valued survival variable X has a log normal distribution and we write $X \sim lognormal(\mu, \sigma)$ if $Y = \ln(X) \sim N(\mu, \sigma)$.

Suppose Φ is the cumulative distribution function of the standard normal random variable, the survival function of $X \sim lognormal(\mu, \sigma)$ is

$$S(x) = P(X > x) = P(\ln(X) > \ln(x)) = 1 - \Phi(\frac{\ln(x) - \mu}{\sigma}),$$

from which the density and hazard may be obtained by differentiating $S(x)$.

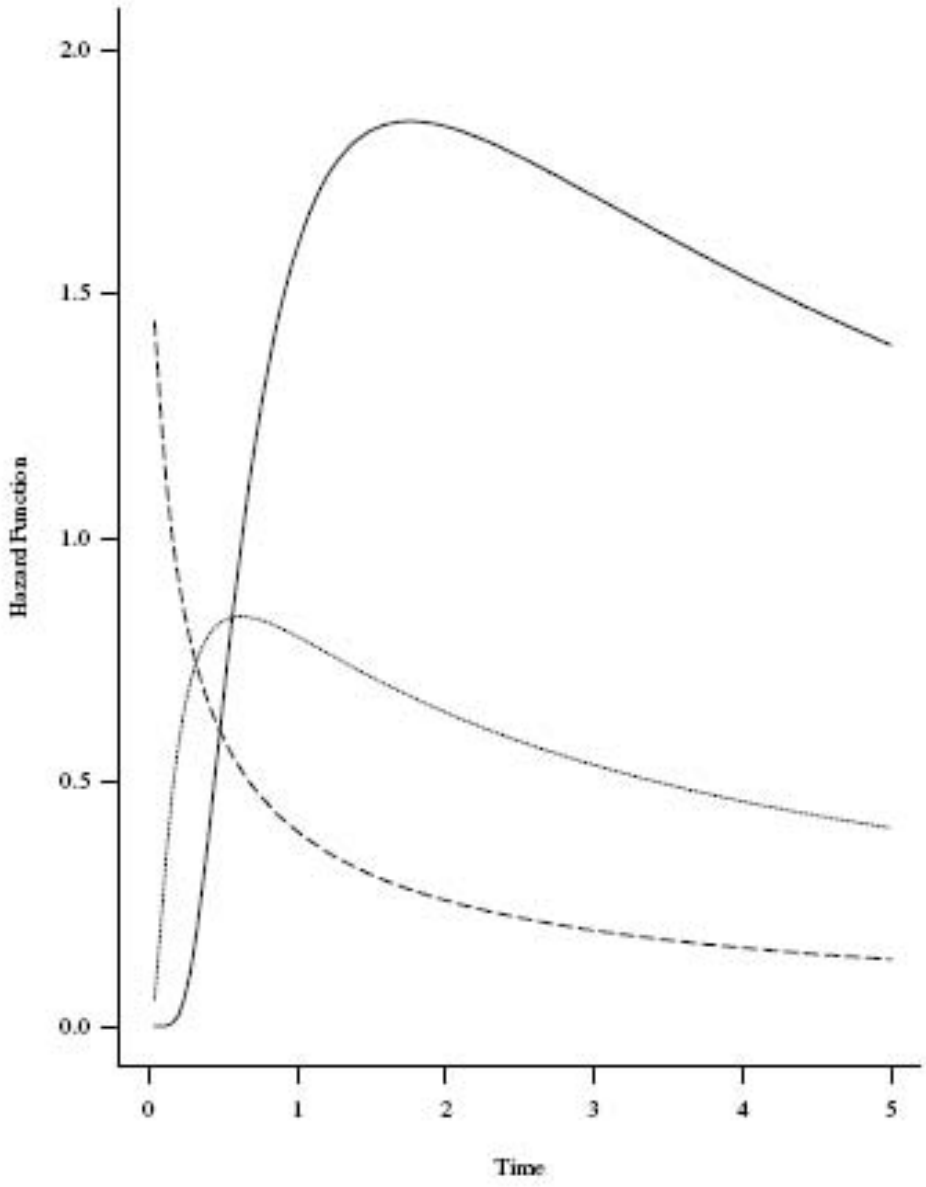
The Log Normal Distribution (2)

The features of the hazard rate: the hazard function of the lognormal is hump-shaped. It initially increases, reaches a maximum and then decreases toward 0 as lifetimes become larger and larger.

- The model is not suitable for lifetime modeling where hazards increase with old age.
- By using a part of the distribution, we can model the onset of some disease.

For the log normal distribution, the mean lifetime is given by $exp(\mu + \sigma^2)$ and the variance by $[exp(\sigma^2) - 1]exp(2\mu + \sigma^2)$.

Lognormal Hazard Function



Log Logistic Distribution

Log-logistic distribution: mimic lognormal distribution properties, but it has closed-form hazard and survival functions.

A positive valued survival variable X has a log logistic distribution and we write $X \sim \mathit{loglogistic}(\mu, \sigma)$ with two parameters μ and σ if $Y = \ln(X)$ follows a logistic distribution with survival function

$$S_Y(y) = 1 - \frac{1}{1 + exp[-(\frac{y-\mu}{\sigma})]}.$$

Log Logistic Distribution (2)

The survival function of the log logistic distribution is

$$S_X(x) = 1 - \frac{1}{1 + exp[-(\frac{\ln(x)-\mu}{\sigma})]} = \frac{1}{1+\lambda x^{\alpha}},$$

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

The hazard function is

$$h_X(x) = \frac{\alpha\lambda x^{\alpha}}{(1+\lambda x^{\alpha})^2},$$

which is similar in shape to the log normal hazard, but it is considerably easier to manipulate.

Regression Models

- To adjust the survival function to account for the covariate/explanatory variables
- Examples:
  - Quantitative variables: blood pressure, temperature, age, weight
  - Qualitative variables: gender, race, treatments, disease status
- The explanatory variable is denoted $\boldsymbol{z} = (z_1, \cdots, z_p)^t$
- Two approaches in regression:
  - Modeling $Y = \ln(X)$, accelerated failure time model
  - Modeling the hazard function $h(x)$
    - Multiplicative hazard rate models
    - Additive hazard rate models

Accelerated Failure Time Model

- The first approach is analogous to the classical linear regression approach
- $Y = \ln(X) = \mu + \gamma^t \boldsymbol{z} + \sigma W$ where γ is the regression coefficient vector, W is the error distribution
- If the error distribution is normal, then the regression model is a lognormal
- If the error distribution is the smallest extreme value distribution, then the regression model is the Weibull
- This model is called accelerated failure time model
- Let $S_0(x)$ denote the sf of X when $\boldsymbol{z} = 0$, then $S(x|\boldsymbol{z}) = S_0[x \exp(-\gamma^t \boldsymbol{z})]$
  - If $\exp(-\gamma^t \boldsymbol{z}) > 1$ the time scale is accelerated
  - If $\exp(-\gamma^t \boldsymbol{z}) < 1$ the time scale is decelerated

Example

- Assume X is a Weibull distribution, $Y = \ln(X)$
- The regression model is $Y = \gamma^t \boldsymbol{z} + \sigma W$
- The sf under baseline is $S_0(x) = \exp(-x^\alpha)$
- The sf under $\boldsymbol{z}$ is
  $S(x|\boldsymbol{z}) = \exp\{-[x \exp(-\gamma^t \boldsymbol{z})]^\alpha\} = S_0[x \exp(-\gamma^t \boldsymbol{z})]$

Multiplicative Hazard Model

- The Second approach is to model the hazard function as a function of covariate
- The first class of model in this approach is called the multiplicative hazard rate model
- In particular, the conditional hazard rate of an individual with covariate $\boldsymbol{z}$ is a product of the baseline hazard and a non-negative function of the covariate $c(\beta^t \boldsymbol{z})$, that is $h(x|\boldsymbol{z}) = h_0(x)c(\beta^t \boldsymbol{z})$
- For Cox model, $c(\beta^t \boldsymbol{z}) = \exp(\beta^t \boldsymbol{z})$
- A key feature of this class of model is proportional hazard when $\boldsymbol{z}$ is fixed $\frac{h(x|\boldsymbol{z}_1)}{h(x|\boldsymbol{z}_2)} = \frac{h_0(x)c(\beta^t \boldsymbol{z}_1)}{h_0(x)c(\beta^t \boldsymbol{z}_2)} = \frac{c(\beta^t \boldsymbol{z}_1)}{c(\beta^t \boldsymbol{z}_2)}$
- The sf under $\boldsymbol{z}$ is $S(x|\boldsymbol{z}) = S_0(x)^{c(\beta^t \boldsymbol{z})}$

Example

- Consider the baseline hazard to be a Weibull, that is $h_0(x) = \alpha\lambda x^{\alpha-1}$
- $h(x|\boldsymbol{z}) = \alpha\lambda x^{\alpha-1} c(\beta^t \boldsymbol{z})$
- If Cox model is used $h(x|\boldsymbol{z}) = \alpha\lambda x^{\alpha-1} \exp(\beta^t \boldsymbol{z})$
- The sf under $\boldsymbol{z}$ is $S(x|\boldsymbol{z}) = \exp[-\lambda x^{\alpha}]^{\exp(\beta^t \boldsymbol{z})} = \exp[-\lambda(x \exp[\beta^t \boldsymbol{z}/\alpha])^{\alpha}]$
- This is in the form of the accelerated failure time model
- The Weibull is the only continuous distribution which has the property of being both an accelerated failure time and a multiplicative hazards model

Additive Hazard Model

- The second class of models for the hazard rate is the family of additive hazard rate model
- The condition hazard function is modeled by

$$h(x|\boldsymbol{z}) = h_0(x) + \sum_{j=1}^{p} z_j(x)\beta_j(x)$$

- More details are in Chapter 10 (will not cover)