SOC 7717 Event History Analysis and Sequence Analysis

Week 5: Parametric Models

Wen Fan

Spring 2019

Outline

Overview

A mathematical framework for survival analysis

The aim: to derive a mathematical representation of the relationship between a response variable and a number of explanatory variables, together with a measure of the uncertainty of any such relationship

- Parametric distributions can be expressed as a mathematical function of the parameters
  - E.g., the binomial probability distribution is a function of two parameters—n and p: $\Pr(X = r) = \frac{n!}{r!(n-r)!} p^r (1-p)^{n-r}$ for $r = 0, 1, ..., n$
- A nonparametric distribution is one which cannot be expressed as a mathematical function

Parametric models

If we assume that survival times follow a given distribution, we could model the survival time or the hazard as a function of one or more covariates

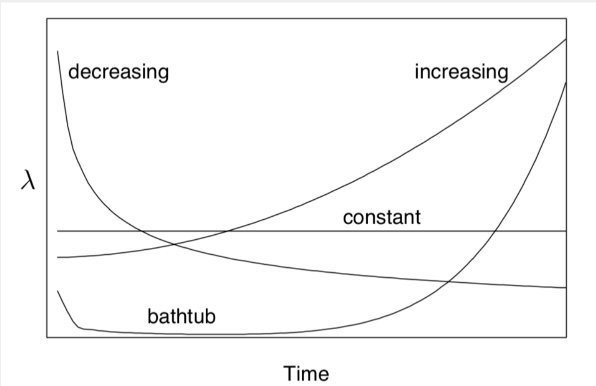
Two classes of parametric models in survival analysis

- Accelerated failure time models: log-time parameterization
- Proportional hazards models: hazard parameterization

Advantages of parametric models

- Accommodates right censoring, left censoring, and interval censoring
- Allows us to test certain hypotheses about the shape of the hazard function
- More efficient estimates (if the shape of the survival distribution is known)
- Can easily generate predicted event times and hazard rates

Common shapes of the hazard function



Accelerated Failure Time Models and Proportional Hazards Models

An Example

Let's first suppose that there is no censoring

- Sample: 100 females, interviewed at age 45
- T = age at first marriage or censoring time
- $X1$ = race
- $X2$ = father's years of schooling

Accelerated Failure Time (AFT) models

Let ϵ be a random disturbance, we can write the survival time as:

$$\log T_i = \beta_0 + \beta_1 X_i 1 + \beta_2 X_i 2 + \sigma \epsilon_i,$$

where $\beta_0, \beta_1, \beta_2,$ and σ are parameters to be estimated.

- This is a general class of models, which differ depending on distribution of ϵ_i
- $var(\epsilon_i)$ is fixed, but the parameter σ allows the disturbance variance to vary
- Why log? $T > 0$ for all values of X

AFT models

$$\log T_i = \beta_0 + \beta_1 X_i 1 + \beta_2 X_i 2 + \sigma \epsilon_i$$

The only differences between this model and the usual linear regression model are that there is a σ before the ϵ and the dependent variable is logged

- The σ can be omitted, but it's simpler to fix the variance of ϵ at some standard values (e.g., 1) and let σ change in value to accommodate changes in the disturbance variance
- If we assume T has a log-normal distribution, then $\log(T)$ has a normal distribution $\longrightarrow$ We can readily estimate this model by OLS

Alternative distributions of ϵ and T

Distribution of ϵ	Distribution of T (conditional on X's)
Normal	Log-normal
Logistic	Log-logistic
Extreme value	Weibull (or exponential, a special case)
Log-gamma	Gamma

Note that all AFT models are named for the distribution of T rather than for the distribution of ϵ

These alternatives lead to different hazard functions and thus different substantive interpretations

Proportional hazards (PH) models

The AFT models can also be expressed in terms of hazard functions $\longrightarrow$ proportional hazards (PH) models:

$$h(t) = h_0(t)\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k),$$

where $h_0(t)$, the baseline hazard, can take different functions and usually involves interaction with time

The coefficients can be interpreted as relative hazard ratios

Common Parametric Models in Survival Analysis

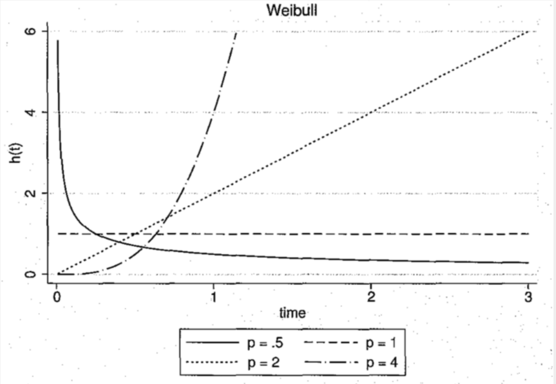
1. Exponential models

- T has an exponential distribution
- The AFT model is $\log T = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + \epsilon$, where ϵ has a standard extreme value distribution
- The PH model is $h(t) = \exp(\beta_0^\star + \beta_1^\star X_1 + \dots + \beta_k^\star X_k)$
  - $\beta_j = -\beta_j^\star$
- Constant hazard function (memoryless): useful as a baseline model, simplifies calculations, but may not be realistic
- Can be very useful for discrete data (piecewise exponential model, to be discussed later)

2. Weibull models

- T has a Weibull distribution
- The AFT model is $\log T = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + \sigma\epsilon$, where ϵ has a standard extreme value distribution and σ can take any positive values
- The PH model is $h(t) = pt^{p-1}\exp(\beta_0^\star + \beta_1^\star X_1 + \dots + \beta_k^\star X_k)$
  - When $t = 0$, $h(t) = 0$ or ∞
  - $p < 1$: decreasing hazard
  - $p > 1$: increasing hazard
  - $p = 1$: constant hazard (exponential model)
  - $\beta_j = \frac{-\beta_j^\star}{p}$

Hazard functions for Weibull models



2. Weibull models

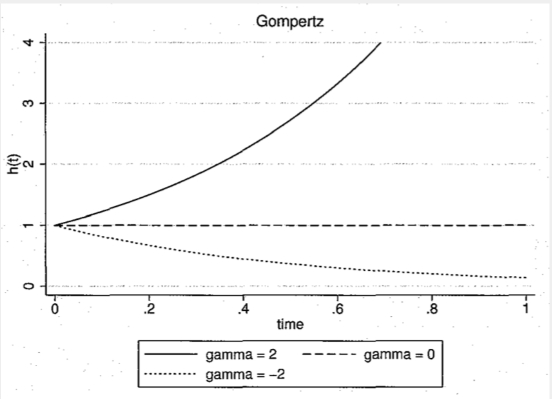
A very popular parametric model

- Has a relatively simple survivor function that is easy to manipulate mathematically
- Both an AFT model and a PH model

3. Gompertz models

- T has a Gompertz distribution
- The PH model is $h(t) = \exp(\gamma t) \exp(\beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k)$
  - $\gamma < 0$: hazard decrease with time
  - $\gamma > 0$: hazard increases with time
  - $\gamma = 0$: constant hazard (exponential model)
- It's not an AFT model
- Suitable for modeling data with monotone hazard rates that either increase or decrease exponentially with time

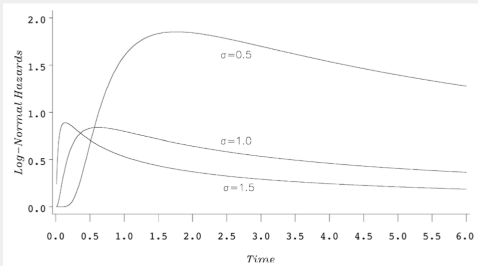
Hazard functions for Gompertz models



4. Log-normal models

Non-monotonic hazard function: $h(t) = 0$ when $t = 0$, increases to a maximum, then decreases asymptotically to 0 as t goes to infinity $\longrightarrow$ the inverted U-shape is appropriate for repeatable events

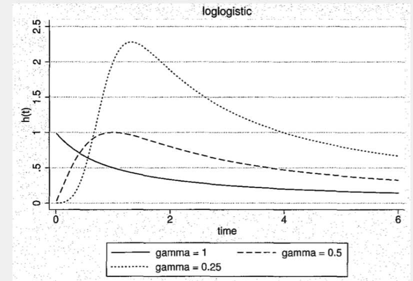
It's not a PH model, so no closed-form hazard function



5. Log-logistic models

- If $\gamma < 1$, $h(t)$ behaves like log-normal models
- If $\gamma \geq 1$, $h(t)$ is monotone decreasing

It's not a PH model, so no closed-form hazard function



6. Generalized gamma models

A generalized model with three parameters (β_0, κ, σ): the exponential ($\kappa = \sigma = 1$), Weibull ($\kappa = 1$), and log-normal ($\kappa = 0$) models are special cases of the Gamma model

In addition, for some values of the parameters, the Gamma can have a "bathtub" shape

Complicated formula, long computation time, convergence problem

Comparison of models

Distribution	Metric	Hazard Shape
Exponential	AFT, PH	constant
Weibull	AFT, PH	monotone
Gompertz	PH	monotone
Log-normal	AFT	variable
Log-logistic	AFT	variable
Generalized Gamma	AFT	variable

Estimation of the AFT models

- When there is no censoring, can use OLS regression (with $\log(T)$ as dependent variable)
- An alternative is maximum likelihood
  - Equivalent to OLS for normal data. More efficient for other models
  - Need iterative methods for non-normal distributions
  - Good large-sample properties: consistent, asymptotically efficient, asymptotically normal
  - Straightforward to derive MLEs when there are no other obvious possibilities (e.g., presence of censoring)

Likelihood construction

- Exact event time: $T_i = t_i \longrightarrow L_i = f(t_i)$
- Right censored: $T_i > t_i \longrightarrow L_i = S(t_i)$
- Left censored: $T_i \leq t_i \longrightarrow L_i = 1 - S(t_i)$
- Interval censored: $T_i \in (B_i, U_i] \longrightarrow L_i = S(B_i) - S(U_i)$
- Left truncated: $(y_i, t_i) \longrightarrow L_i = \frac{f(x_i)}{S(y_i)}$

Parametric models in Stata

```
streg varlist, dist(distname) nohr time tr nolog
```

- `dist(distname)` specifies the survival model to be estimated (choose from `exponential`, `weibull`, `gompertz`, `lognormal`, `loglogistic`, and `ggamma`)
- To obtain unexponentiated coefficients for PH models, use the `nohr` option
- To fit AFT models, use the `time` option
- To obtain exponentiated coefficients for AFT models, use the `tr` option, which stands for "time ratio"

Types of Estimates Available for Various Models

	Log-normal, Log-logistic, Gamma	Weibull, Exponential	Gompertz
AFT			
β	default	`time`	NA
$\exp(\beta)$	`tr`	`tr`	NA
PH			
β	NA	`nohr`	`nohr`
$\exp(\beta)$	NA	default	default

Interpretation and Presentation of Estimates

Interpretation of coefficients

Suppose we have an AFT model. To get the percentage change in survival time for a 1-unit increase in X, we calculate

$$100(e^{\beta} - 1)$$

The same principle applies to PH formulation, except that percentage increase in survival times correspond to percentage decreases in the hazard

Generating predicted survival times

After fitting a model with `streg`, it is easy to get predicted times to events by using the `predict` command

- To get predicted median times, use:
  - » `predict var, median time`
- To get predicted mean times to event, use:
  - » `predict var, mean time`

Marginal effects

Graphing the AFT hazard function

For any model you fit with `streg`, the `stcurve` command can be used to produce a graph of the hazard function that is implied by the fitted model

Hypothesis Testing and Model Comparisons

Methods for hypothesis testing

- Wald test
  - Parameter estimates and their estimated variances/covariances
- Likelihood-ratio test
  - Log-likelihoods
- Score test
  - First and second derivatives of the log-likelihood function

Methods for hypothesis testing

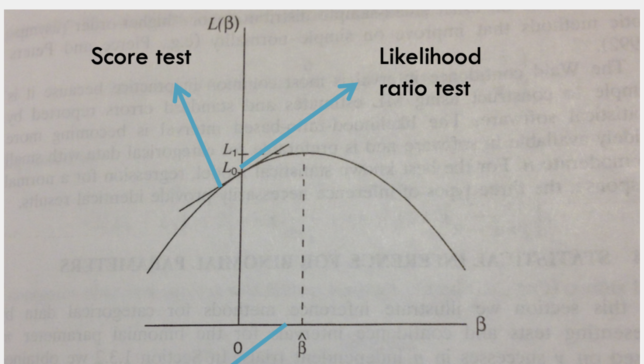


FIGURE 1.1 Log-likelihood function and information used in three tests of H_0: $\beta = 0$.

Source: Alan Agresti *Categorical Data Analysis* (p.13).

Comparing models

Nested models

- Wald test or likelihood-ratio test
- The exponential model is nested within the Weibull model, while the Weibull and log-normal models are nested within the generalized Gamma model

Non-nested models

- Information criterion statistics (AIC, BIC): can compare nested and non-nested models
- Think of them as models of fit penalized based on sample size and number of parameters

Choice of Origin for Measurement of Time

Choice of origin for measurement of time

For parametric models, the decision regarding the origin of time is vital

But how do we decide?

Obvious in many cases:

- Recidivism: release from prison
- Marital dissolution: date of marriage
- Job termination: start of job

Not so obvious in other cases:

- Retirement: age or time in the labor force?
- First birth: age or time since marriage began?
- Death due to cancer: age or time since diagnosis?

Choice of origin for measurement of time

Must choose one as the principal time axis, although it may be possible to include others as explanatory variables

Some possibilities:

- Age
- Calendar time
- Time since last event of the same type
- Time since some other relevant event

Choice of origin for measurement of time

Criteria for choice:

- Choose the beginning of continuous exposure to risk
- In experimental studies, choose the time of randomization to treatment as the time origin
- Choose the time origin that has the strongest effect on the hazard

An Introduction to Likelihood Inferences

Likelihood inference

- The aim of statistical inference is to estimate population parameters of interest from observed data
- Imagine we are interested in estimating the proportion, p, that a toss of a coin will result in heads
- Data: We toss the coin 10 times and observe 4 heads

Likelihood inference

- We wish to estimate the parameter of interest, p, from the observed data (the 10 tosses of the coin). Issues of interest may be:
  - What is the most likely value for p?
  - What is a range of likely values for p?

The likelihood approach is to calculate the probability of observing the observed data, given the probability model, for all possible values of the parameter(s) of interest and choosing the values of the parameter(s) that make the data most likely

Likelihood inference

- In other words, for what value of p is the probability of tossing 4/10 heads most likely?
- If the true value is $p = 0$, what is the probability of observing 4 heads in 10 tosses?
- That one was easy, but what if $p = 0.1$?
- If $p = 0.1$, the number of observed heads can theoretically be any integer between 0 and 10 and the probability of each is described by the binomial distribution

Likelihood inference

- Recall that if X is a random variable described by a binomial distribution with parameters n and p, the probability distribution of X is given by:

$$\Pr(X = r) = \frac{n!}{r!(n-r)!} p^r (1-p)^{n-r}$$

- $Pr(X = r)$ is the probability of obtaining r "successes" (e.g., toss heads) in a sample of size n where the true proportion is p

Likelihood inference

For $p = 0.1$ and $n = 10$ the probability of each of the possible outcomes is:

r	Prob(r heads)
0	0.35
1	0.39
2	0.19
3	0.06
4	0.01
5	0.00
6	0.00
7	0.00
8	0.00
9	0.00
0	0.00
10	0.00
Σ	1.00

Likelihood for a range of values of p

p	Prob(r = 4)
0.00	0.00
0.10	0.01
0.20	0.09
0.30	0.20
0.40	0.25
0.50	0.21
0.60	0.11
0.70	0.04
0.80	0.01
0.90	0.00
1.00	0.00
Σ	1.00

This is the likelihood function for p ranging from 0.00 to 1.00

What are likely values for p?

- The value of p for which the likelihood is greatest is $p = 0.4$. This is called the maximum likelihood estimate (MLE)
- We can see that $p = 0.5$ is also quite likely. The probability of the data is 0.21 when $p = 0.5$, not that different from a probability of 0.25 when $p = 0.4$ (the MLE)
- We can test whether $p = 0.5$ is a likely value by studying the ratio of the likelihoods: $L(0.5)/L(0.4) = 0.21/0.25 = 0.8176$

What are likely values for p?

A result in mathematical statistics tells us that, if the true value of p was 0.5, then minus twice the log likelihood ratio will have a chi-square distribution with 1 degree of freedom

$$-2 \ln[L(0.5)/L(0.4)] = -2[l(0.5) - l(0.4)] = 0.403$$

The chi-square value at 1 degree of freedom and 5 percent probability is 3.84. What's your conclusion?

Mathematically

We wish to find the value of p that maximizes the likelihood function

$$L(p) = \frac{n!}{r!(n-r)!} p^r (1-p)^{n-r}$$

It is generally easier to maximize the log likelihood (the maximum will occur at the same value). Ignoring the constant,

$$l(p) = \ln[L(p)] = r \log(p) + (n-r) \log(1-p)$$

- The derivative of $l(p)$ with respect to p is $l'(p) = \frac{r}{p} - \frac{(n-r)}{(1-p)}$
- The maximum value of $l(p)$ will occur when $l'(p) = 0$, which is $\hat{p} = \frac{r}{n}$