Survival Analysis

Lucila Ohno-Machado
Brigham and Women’s Hospital

Outline
- Basic concepts & distributions
  - Censored data
  - Survival, Hazard
  - Parametric models
  - Non-parametric models
- Univariate models
  - Life-table
  - Product-Limit
- Multivariate models
  - Cox proportional hazard
  - Neural nets

What we are trying to do
- Predict survival
- and evaluate performance on new cases
- and determine which variables are important

Censoring
- Type I: Non-censored
- Type II: Left-censored

Survival function
- Probability that an individual survives at least t
  - \( S(t) = P(T > t) \)
  - By definition, \( S(0) = 1 \) and \( S(\infty) = 0 \)
  - Estimated by (# survivors at \( t \) / total patients)

Probability density function
- \( f(t) = \lim_{\Delta t \to 0} \frac{P(\text{individual dies (t,t+\Delta t)})}{\Delta t} \)
- \( f(t) \) always non-negative
- Area below density is 1
- Unconditional failure rate
- Estimated by # patients dying in the interval/(total patients*interval_width)
Hazard function

- \( h(t) = \lim_{\Delta t \to 0} \frac{P(\text{individual aged } t \text{ dies } (t, t+\Delta t))/\Delta t }{\Delta t} \)
- \( h(t) \) is instantaneous failure rate
- Conditional failure rate
- Estimated by 
  \( \frac{\text{# patients dying in the interval}}{\text{survivors at } t \times \text{interval width}} \)

Some other definitions and derived relationships

- \( S(t) \) is "cumulative" survival
- \( F(t) \) is cumulative probability
- \( S(t) = 1 - F(t) \)
- \( H(t) = -\log S(t) \)
- \( h(t) \) is cumulative hazard

- \( h(t) = f(t)/S(t) \)
- \( f(t) = -S'(t) \)
- \( f(t) = h(t) e^{-H(t)} \)
- \( S(t) = e^{-\lambda t} \)
- \( H(t) = \lambda t \)
- \( h(t) = \lambda S(t) = \lambda e^{-\lambda t} \)

- For \( \gamma > 0 \)
  - \( f(t) = \gamma t^{\gamma-1} e^{-\gamma t} \)
  - \( S(t) = e^{-\gamma t} \)
  - \( h(t) = \gamma t^{\gamma-1} e^{-\gamma t} \)

- \( S(t) = 1 - \sum_{j=1}^{\infty} \frac{d_j}{n_j} \prod_{i=1}^{j-1} \left(1 - \frac{d_i}{n_i}\right) \)

- \( d_j \) is the number of deaths in interval \( j \)
- \( n_j \) is the number of individuals at risk
- Product is from time interval 1 to \( j \)
- One interval per death time

Weibull distribution

- Generalization of the exponential
- Example: \( \gamma > 0 \)
  - \( f(t) = \gamma t^{\gamma-1} e^{-\gamma t} \)
  - \( S(t) = e^{-\gamma t} \)
  - \( h(t) = \gamma t^{\gamma-1} e^{-\gamma t} \)
- Example: \( \gamma = 2 \)
- Example: \( \gamma = 1 \)

Kaplan-Meier

- Example
- Deaths: 10, 37, 40, 80, 91, 143, 164, 188, 188, 190, 192, 206, …
Life-Tables

- AKA actuarial method
  \[ S(t) = \prod (n_j - d_j)/n_j \]
  - \( d_j \) is the number of deaths in interval \( j \)
  - \( n_j \) is the number of individuals at risk
  - Product is from time interval 1 to \( j \)
- Pre-defined intervals \( j \) are independent of death times

Life-Table

Univariate models

Mantel-Haenzel test (log-rank test)

- Compares two survival curves that obey the "power law" relation between survival curves
  \[ S_1(t) = [S_j(t)]^{\Theta_{ij}} \]
  where \( \Theta_{ij} \) is independent of time
- Example: \[ S_1(t) = e^{-\lambda t} \]
  \[ S_2(t) = e^{-k \lambda t} \]

Multiple strata

2 x 2 table for 1 time point

- \( P(d_1) \) has hypergeometric distribution
  \[ X^2 = \frac{(d_1 - E(d_1))^2}{\text{Var}(d_1)} \]
  \( \chi^2 \) with 1 d.f.

- \( E(d_1) = n_1 (D/N) \)
- \( \text{Var}(d_1) = n_1 (D/N) (1-D/N) [(N - n_1)/(N-1)] \)
2 x 2 tables for all time points (in which death occurred)

\[
\sum_{j=1}^{k} \frac{(d_{j1} - E(d_{j1}))^2}{\text{Var}(d_{j1})}
\]

\( \chi^2 \) with 1 d.f.

- \( E(d_{j1}) = n_{j1} (D_j/N_j) \)
- \( \text{Var}(d_{j1}) = E(d_{j1}) n_{j2} (N_j - D_j) / [N_j(N_j-1)] \)

Proportional Hazards

\[ \lambda_i = \lambda e^{-|x_i|} \]

where \( \lambda \) is baseline hazard and \( x_i \) is covariate for patient

Cox proportional hazards

\[ h_i(t) = h_0(t) e^{\beta x_i} \]

- Survival

\[ S_i(t) = \left[ S_0(t) \right] e^{\beta x_i} \]

Multivariate models

- Several strata, each defined by a set of variable values
- Can potentially go as far as “one stratum per case”
- Prediction for individuals based on several variables

Proportional Hazards

\[ h_i(t) = h_0(t) e^{\beta x_i} \quad S_i(t) = \left[ S_0(t) \right] e^{\beta x_i} \]

when \( x_i \) is a binary variable

\[ S_2(t) = \left[ S_1(t) \right] e^{\beta} \]

This form of the Cox model leads to the Mantel-Haenszel significance test

Cox Proportional Hazards

\[ h_i(t) = h_0(t) e^{\beta x_i} \]

- New likelihood function is how we estimate \( \beta \)
- From the set of individuals at risk at time \( j \) (\( R_j \)), the probability of picking exactly the one who died is

\[ \frac{h_i(t) e^{\beta x_i}}{\sum_m h_m(t) e^{\beta x_m}} \]

- Then likelihood function to maximize to all \( j \) is

\[ L(\beta) = \prod \left( e^{\beta x_i} / \sum_m e^{\beta x_m} \right) \]

Important details

- Survival curves can't cross (hazards are proportional)
- There is a common baseline \( h_0 \), but we don't need to know it to estimate the coefficients
- We don't need to know the shape of hazard function
- Cox model is commonly used to interpret importance of covariates (amenable to variable selection methods)
- It is the most popular multivariate model for survival
- Testing the proportionality assumption is difficult and hardly ever done
Estimating survival for a patient with the Cox model

- Need to estimate the baseline
- Can use parametric or non-parametric model to estimate the baseline
- Can then create a continuous "survival curve estimate" for a patient
- Baseline survival can be, for example:
  - the survival for a case in which all covariates are set to their means
  - Kaplan-Meier estimate for all cases

Example of survival estimates

What if the proportionality assumption is not OK?

- Survival curves may cross
- Other multivariate models can be built
- Survival at certain time points are modeled and combined

Example of survival estimates

Single-point models

- Logistic regression
- Neural nets

Problems

- Dependency between intervals is not modeled (no links between networks)
- Nonmonotonic curves may appear
- How to evaluate?

Table III. Calibration of NN models. Results of Hosmer-Lemeshow test. The p value (p) assesses whether there is a significant difference (at a significance level p) between the number of predicted deaths and the actual number of patients who died within certain years of AIDS diagnosis.

<table>
<thead>
<tr>
<th>Year of follow-up</th>
<th>p</th>
<th>p*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7.0256</td>
<td>0.1045</td>
</tr>
<tr>
<td>2</td>
<td>6.4206</td>
<td>0.1132</td>
</tr>
<tr>
<td>3</td>
<td>15.7892</td>
<td>0.0024</td>
</tr>
<tr>
<td>4</td>
<td>2.2892</td>
<td>0.1180</td>
</tr>
<tr>
<td>5</td>
<td>15.3148</td>
<td>0.0107</td>
</tr>
<tr>
<td>6</td>
<td>11.3566</td>
<td>0.3728</td>
</tr>
</tbody>
</table>

Table IV. Rejection of isolated NN modules

<table>
<thead>
<tr>
<th>Year of follow-up</th>
<th>Glutenoma cases outside the ROC curves</th>
<th>Adjacent cases outside the ROC curves</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.7154</td>
<td>0.6913</td>
</tr>
<tr>
<td>2</td>
<td>0.7379</td>
<td>0.7185</td>
</tr>
<tr>
<td>3</td>
<td>0.7398</td>
<td>0.7213</td>
</tr>
<tr>
<td>4</td>
<td>0.7353</td>
<td>0.6979</td>
</tr>
<tr>
<td>5</td>
<td>0.6547</td>
<td>0.6398</td>
</tr>
<tr>
<td>6</td>
<td>0.6546</td>
<td>0.7373</td>
</tr>
</tbody>
</table>

*Using k-fold cross validation (10 folds for each model).
Accounting for dependencies

- "Link" networks in some way to account for dependencies

<table>
<thead>
<tr>
<th>Year</th>
<th>Survival (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.3</td>
</tr>
<tr>
<td>4</td>
<td>0.2</td>
</tr>
</tbody>
</table>

Table V. Calibration of combined NN models

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0129</td>
<td>0.0995</td>
<td>0.0593</td>
</tr>
<tr>
<td>0.0129</td>
<td>0.0995</td>
<td>0.0593</td>
</tr>
</tbody>
</table>

Table VI. Resolution for combined NNs

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area under ROC</td>
<td>Area under ROC</td>
<td>Area under ROC</td>
</tr>
<tr>
<td>ROC</td>
<td>ROC</td>
<td>ROC</td>
</tr>
</tbody>
</table>

Survival without Coronary Disease

<table>
<thead>
<tr>
<th>Year of follow-up</th>
<th>Test set</th>
<th>Area under the ROC curve</th>
<th>Standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.0118</td>
<td>0.7038</td>
<td>0.0242</td>
</tr>
<tr>
<td>4</td>
<td>0.1339</td>
<td>0.7147</td>
<td>0.0190</td>
</tr>
<tr>
<td>6</td>
<td>0.6175</td>
<td>0.7255</td>
<td>0.0152</td>
</tr>
<tr>
<td>8</td>
<td>0.3247</td>
<td>0.7337</td>
<td>0.0138</td>
</tr>
<tr>
<td>10</td>
<td>0.2653</td>
<td>0.7337</td>
<td>0.0130</td>
</tr>
<tr>
<td>12</td>
<td>0.6464</td>
<td>0.7445</td>
<td>0.0123</td>
</tr>
<tr>
<td>14</td>
<td>0.9273</td>
<td>0.7752</td>
<td>0.0121</td>
</tr>
<tr>
<td>16</td>
<td>0.7588</td>
<td>0.8059</td>
<td>0.0219</td>
</tr>
<tr>
<td>18</td>
<td>0.7588</td>
<td>0.8275</td>
<td>0.0122</td>
</tr>
<tr>
<td>20</td>
<td>0.2278</td>
<td>0.8374</td>
<td>0.0123</td>
</tr>
<tr>
<td>22</td>
<td>1.2683</td>
<td>0.8724</td>
<td>0.0165</td>
</tr>
</tbody>
</table>
Summary

- Kaplan-Meier for univariate
- Cox Proportional for multivariate if survival curves don’t cross
- Other methods for multivariate survival exist: logistic regression, neural nets, CART, etc.