Survival Analysis

STAT 526
Professor Olga Vitek

May 4, 2011
Survival Data and Survival Functions

- Statistical analysis of *time-to-event data*
  - Lifetime of machines and/or parts (called *failure time analysis* in engineering)
  - Time to default on bonds or credit card (called *duration analysis* in economics)
  - Patients survival time under different treatment (called *survival analysis* in clinical trial)
  - *Event-history analysis* (sociology)

- Why a special topic on survival analysis?
  - Non-normal and skewed distribution
  - Need to answer questions related to \( P(X > t + t_0 | X \geq t_0) \)
  - Censored/truncated data
  - Here we only focus on right-censored data
Survival Function

• Continuous survival time $T$
  – its probability density function is $f(t)$
  – its cumulative probability function is $F(t)$
    \[ F(t) = P(T \leq t) = \int_0^t f(s)ds \]

• The survival function of $T$ is
  \[ S(t) = P(T > t) = 1 - F(t) \]
  – also called survival rate
  – steeper curve $\rightarrow$ shorter survival
    \[ S'(t) = -f(t) \]

• Mean survival time
  \[
  \mu = E\{T\} = \int_0^\infty t f(t) \, dt = \int_0^\infty t \, dF(t) = \int_0^\infty t \, d[1 - S(t)]
  = \int_0^\infty \left[ \int_0^t dx \right] d[1 - S(t)] = \int_0^\infty \left[ \int_x^\infty d[1 - S(t)] \right] \, dx
  = \int_0^\infty S(x) \, dx
  \]

9-2
Hazard Function

• The hazard function of $T$ is

$$\lambda(t) = \lim_{\Delta t \downarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$$

- proportion of subjects with the event per unit time, around time $t$; $\lambda(t) \geq 0$
- measure of 'proneness' to the even as function of time
- $\lambda(t) \neq f(t)$: $\lambda(t)$ is conditional on survival to $t$

• Relates to the survival function

$$\lambda(t) = \lim_{\Delta t \downarrow 0} \frac{F(t + \Delta t) - F(t)}{\Delta t} / S(t) = \frac{f(t)}{S(t)} = -\frac{S'(t)}{S(t)}$$

$$= -\frac{d}{dt} \log S(t) \text{ (negative slope of log-survival)}$$

• The cumulative hazard function of $T$ is defined as $\Lambda(t) = \int_0^t \lambda(s)ds$

- $\Lambda(t) = -\log S(t)$
- $S(t) = \exp\{-\Lambda(t)\}$
- $\lambda(t), \Lambda(t)$ or $S(t)$ define the distribution
Parametric Survival Models

• Exponential distribution:
  \[\lambda(t) = \rho, \text{ where } \rho > 0 \text{ is a constant, and } t > 0\]
  \[S(t) = e^{-\rho t}, \Rightarrow f(t) = -S'(t) = \rho e^{-\rho t}\]

• Weibull distribution:
  \[\lambda(t) = \lambda p(\lambda t)^{p-1}; \lambda, p > 0 \text{ are constants, } t > 0\]
  \[S(t) = e^{-(\lambda t)^p}, \Rightarrow f(t) = -S'(t) = e^{-(\lambda t)^p}\lambda p(\lambda t)^{p-1}\]
  \(- \text{ is the Exponential distribution when } p = 1\)

• Gompertz distribution:
  \[\lambda(t) = \alpha e^{\beta t}; \alpha, \beta > 0 \text{ are constants, } t > 0\]
  \(- \text{ is the Exponential distribution when } \beta = 0\)

• Gompertz-Makeham distribution:
  \[\lambda(t) = \lambda + \alpha e^{\beta t}; \alpha, \beta > 0 \text{ are constants, } t > 0\]
  \(- \text{ adds an initial fixed component to the hazard}\)
Non-Parametric Survival Models

- Approximate time by discrete intervals

\[ f(x_i) = \begin{cases} \Pr\{T = x_i\} & \text{otherwise} \\ 0 & \end{cases} \]

\[ S(x_i) = \sum_{j: x_j \geq t} f(x_j) \]

\[ \lambda(x_i) = \frac{f(x_i)}{S(x_i)} \]

- After substitution:

\[ f(x_i) = \lambda(x_i) \prod_{j=1}^{i-1} (1 - \lambda_j) \]

\[ S(x_i) = \prod_{j=1}^{i-1} (1 - \lambda_j) \]

\[ S(t) = \prod_{j: x_j < t} (1 - \lambda_j), \ t > 0 \]
Right-Censored Data

- Survival time of $i$-th subject $T_i$, $i = 1, \cdots, n$.

- Censoring time of $i$-th subject $C_i$, $i = 1, \cdots, n$.

- Observed event for $i$th subject:

\[ Y_i = \min(T_i, C_i), \quad \delta_i = \mathbb{1}_{\{T_i \leq C_i\}} = \begin{cases} 1, & \text{if } T_i \leq C_i \\ 0, & \text{if } T_i > C_i \end{cases} \]

- Data are reported in pairs $(Y_i, \delta_i)$

  $=$ (observedTime, hasEvent)

- Assumption:

  - $C_i$ are predetermined and fixed, or

  - $C_i$ are random, mutually independent, and independent of $T_i$
Non-parametric Estimation of Survival Function And Comparison Between Groups With Censoring
Example: Remission Times of Leukaemia Patients

- 21 leukemia patients treated with drug (\textit{6-mercaptopurine})
- 21 matched controls, no covariates.

```r
> library(MASS)  # Described in Venable & Ripley, Ch.13
> data(gehan)

> gehan
  pair time cens treat
  1   1   1   1 control
  2   1  10   1  6-MP
  3   2  22   1 control
  4   2   7   1  6-MP
  5   3   3   1 control
  6   3  32   0  6-MP
  7   4  12   1 control
  8   4  23   1  6-MP
  9   5   8   1 control
 10  5  22   1  6-MP
 11  6  17   1 control
 12  6   6   1  6-MP
 13  7   2   1 control
 14  7  16   1  6-MP
 15  8  11   1 control
 16  8  34   0  6-MP
 17  9   8   1 control
```

\ldots \ldots
Kaplan-Meier Estimator of Survival Function

- Also called *product limit estimator*
- Re-organize the data

<table>
<thead>
<tr>
<th>Distinct Event Times</th>
<th>$t_1$</th>
<th>$\cdots$</th>
<th>$t_i$</th>
<th>$\cdots$</th>
<th>$t_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td># of Events</td>
<td>$d_1$</td>
<td>$\cdots$</td>
<td>$d_i$</td>
<td>$\cdots$</td>
<td>$d_k$</td>
</tr>
<tr>
<td># Survivors Right Before $t_i$</td>
<td>$n_1$</td>
<td>$\cdots$</td>
<td>$n_i$</td>
<td>$\cdots$</td>
<td>$n_k$</td>
</tr>
</tbody>
</table>

- **Kaplan-Meier estimator** $\hat{S}(t)$:
  - Estimate $P(T > t_1)$ by $\hat{p}_1 = (n_1 - d_1)/n_1$
  - Estimate $P(T > t_i|T > t_{i-1})$ by $\hat{p}_i = (n_i - d_i)/n_i$, $i = 2, \cdots, k$
  - If $t_{i-1} < t < t_i$,
    $$S(t) = P(T > t) = P(T > t_1)P(T > t|T > t_1) = \cdots$$
    $$P(T > t_1)\{ \prod_{m=2}^{i-1} P(T > t_m|T > t_{m-1}) \}P(T > t|T > t_{i-1})$$
  - Estimate $S(t)$ by $\hat{S}(t) = \hat{S}(t_{i-1})\frac{n_i - d_i}{n_i} = \prod_{i:t_i \leq t} \frac{n_i - d_i}{n_i}$

(Kaplan & Meier, 1958, JASA)
Non-Parametric Fit of Survival Curves

- Represent data in the censored survival form
  - create a data structure 'Surv'
  - '+' represents censored time

```r
> library(survival)
> Surv(gehan$time, gehan$cens)
1 10 22 7 3 32+ 12 23 8 22 17 6 2 16
11 34+ 8 32+ 12 25+ 2 11+ 5 20+ 4 19+ 15 6
8 17+ 23 35+ 5 6 11 13 4 9+ 1 6+ 8 10+
```

- K-M estimation of survival function
  - subset of results for the treatment group:

```r
> fit <- survfit(Surv(time, cens) ~ treat, data=gehan)
> summary(fit)

   treat=6-MP
    time n.risk n.event survival std.err lower95%CI upper95%CI
       6    21       3 0.857 0.0764    0.720    1.000
       7    17       1 0.807 0.0869    0.653    0.996
      10    15       1 0.753 0.0963    0.586    0.968
      13    12       1 0.690 0.1068    0.510    0.935
      16    11       1 0.627 0.1141    0.439    0.896
      22     7       1 0.538 0.1282    0.337    0.858
      23     6       1 0.448 0.1346    0.249    0.807
```
Interval K-M Estimator of Survival Function

• Direct interval estimation of $S(t)$
  – The variance of $\hat{S}(t)$ can be estimated by
  \[
  \hat{\text{var}}(\hat{S}(t)) = [\hat{S}(t)]^2 \sum_{i:t_i \leq t} \frac{d_i}{n_i(n_i - d_i)}
  \]
  – This is Greenwood’s formula (Greenwood, 1926)
  – CI for $S(t)$ is $\hat{S}(t) \pm 1.96 \sqrt{\hat{\text{var}}[\hat{S}(t)]}$
  – $S(t) \in [0, 1]$, but $\hat{\text{var}}(\hat{S}(t))$ may be large enough that the usual CI will exceed this interval

• Interval estimation using $\log S(t)$ (preferred)
  – The variance of $\log \hat{S}(t)$ can be estimated as
  \[
  \hat{\text{var}}[\log \hat{S}(t)] = \sum_{i:t_i \leq t} \frac{d_i}{n_i(n_i - d_i)}
  \]
  – Construct the CI for $\log S(t)$
  \[
  [L, U] = \log \hat{S}(t) \pm 1.96 \sqrt{\hat{\text{var}}[\log \hat{S}(t)]},
  \]
  – For CI of $S(t)$, transform the limits with $[e^L, e^U]$.  

9-11
Estimate the Cumulative Hazard Function \( \Lambda(t) = -\log S(t) \)

- Using the Kaplan-Meier estimator of \( \hat{S}(t) \):
  \[
  \hat{\Lambda}(t) = -\log \hat{S}(t)
  \]

- Alternatively, the Nelson-Aalen Estimator:
  \[
  \tilde{\Lambda}(t) = \begin{cases} 
  0, & \text{if } t \leq t_1 \\
  \sum_{i:t_i \leq t} d_i/n_i, & \text{if } t \geq t_i
  \end{cases}
  \]

  \[
  \text{Var}(\tilde{\Lambda}(t)) = \sum_{i:t_i \leq t} \frac{d_i}{n_i^2}
  \]

  - better small-sample-size performance than based on the K-M procedure
  - useful in comparing the fit of a parametric model to its non-parametric alternative
Non-Parametric Fit of Survival Curves

```r
> plot(fit, conf.int=TRUE, lty=3:2, log=TRUE,
   xlab="Remission (weeks)", ylab="Log-Survival", main="Gehan")
```

- Plot curves and CI; default CI are on the log scale
- `log=TRUE` plots y axis (i.e. $S(t)$) on the log scale (i.e. y axis shows the negative cumulative hazard)
Log-Rank Test for Homogeneity

- Non-parametric test
  - Compare two populations with hazard functions $\lambda_i(t), i = 1, 2$.
  - Collect two samples from each population.
  - Construct a pooled sample with $k$ distinct event times

<table>
<thead>
<tr>
<th>Pool Sample</th>
<th>Distinct Failure Time</th>
<th>$t_1$</th>
<th>$\cdots$</th>
<th>$t_i$</th>
<th>$\cdots$</th>
<th>$t_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td># of Failures</td>
<td>$d_1$</td>
<td>$\cdots$</td>
<td>$d_i$</td>
<td>$\cdots$</td>
<td>$d_k$</td>
<td></td>
</tr>
<tr>
<td># survivors right before $t_i$</td>
<td>$n_1$</td>
<td>$\cdots$</td>
<td>$n_i$</td>
<td>$\cdots$</td>
<td>$n_k$</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample 1</th>
<th>Distinct Failure Time</th>
<th>$d_{11}$</th>
<th>$\cdots$</th>
<th>$d_{1i}$</th>
<th>$\cdots$</th>
<th>$d_{1k}$</th>
</tr>
</thead>
<tbody>
<tr>
<td># of Failures</td>
<td>$n_{11}$</td>
<td>$\cdots$</td>
<td>$n_{1i}$</td>
<td>$\cdots$</td>
<td>$n_{1k}$</td>
<td></td>
</tr>
<tr>
<td># survivors $t_i$ right before $t_i$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample 2</th>
<th>Distinct Failure Time</th>
<th>$d_{21}$</th>
<th>$\cdots$</th>
<th>$d_{2i}$</th>
<th>$\cdots$</th>
<th>$d_{2k}$</th>
</tr>
</thead>
<tbody>
<tr>
<td># of Failures</td>
<td>$n_{21}$</td>
<td>$\cdots$</td>
<td>$n_{2i}$</td>
<td>$\cdots$</td>
<td>$n_{2k}$</td>
<td></td>
</tr>
<tr>
<td># survivors right before $t_i$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Note $d_i = d_{1i} + d_{2i}, n_i = n_{1i} + n_{2i}, i = 1, \cdots, k$
- Test the hypotheses

$$H_0 : \lambda_1(t) = \lambda_2(t), t \leq \tau$$

vs.

$$H_a : \lambda_1(t) \neq \lambda_2(t) \text{ for some } t \leq \tau$$
Log-Rank Test for Homogeneity: Procedure

- Consider \([t_i, t_i + \Delta)\) for small \(\Delta\)

<table>
<thead>
<tr>
<th>Sample 1</th>
<th>Sample 2</th>
<th>Pooled Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td># failures</td>
<td>(d_{1i})</td>
<td>(d_{2i})</td>
</tr>
<tr>
<td># survivors at (t_i)</td>
<td>(n_{1i} - d_{1i})</td>
<td>(n_{2i} - d_{2i})</td>
</tr>
<tr>
<td># of survivors (t_i) right before (t_i)</td>
<td>(n_{1i})</td>
<td>(n_{2i})</td>
</tr>
</tbody>
</table>

- If \(\lambda_1(t) \neq \lambda_2(t)\) around \(t_i\), there should be association between sample and event in this \(2 \times 2\) table

- \(d_{1i}(n_i, d_i, n_{1i})^{H_0} \sim \text{Hypergeometric}(n_i, d_i, n_{1i})\):

\[
P(d_{1i} = x | n_i, d_i, n_{1i}) = \binom{d_i}{x} \binom{n_i - d_i}{n_{1i} - x} \binom{n_i}{n_{1i}}
\]

\[
\implies E[d_{1i} | n_i, d_i, n_{1i}] = \frac{d_i n_{1i}}{n_i}
\]

- Define \(U = \sum_{i=1}^{k} \left[ d_{1i} - \frac{d_{1i} n_{1i}}{n_i} \right]\)

\[
E[U] = 0, \quad \hat{\text{var}}(U) = \sum_{i=1}^{k} \frac{d_{i} n_{1i} (n_i - d_i) (n_i - n_{1i})}{n_i^2 (n_i - 1)}
\]

\[
U/\hat{\text{var}}(U)^{1/2} \overset{\text{asy}}{\sim} N(0, 1)
\]
Log-Rank Test in \( \mathbb{R} \)
(Venable & Ripley Sec. 13.2)

```r
> ?survdiff

> survdiff(Surv(time,cens)~treat,data=gehan)

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Observed</th>
<th>Expected</th>
<th>(O-E)^2/E</th>
<th>(O-E)^2/V</th>
</tr>
</thead>
<tbody>
<tr>
<td>treat=6-MP</td>
<td>21</td>
<td>9</td>
<td>19.3</td>
<td>5.46</td>
<td>16.8</td>
</tr>
<tr>
<td>treat=control</td>
<td>21</td>
<td>21</td>
<td>10.7</td>
<td>9.77</td>
<td>16.8</td>
</tr>
</tbody>
</table>

Chisq= 16.8 on 1 degrees of freedom, p= 4.17e-05

- Conclusion: Reject \( H_0 \)

- Warning: this test does not adjust for covariates, and may often be inappropriate
  - Appropriate for this study, since the individuals are matched pairs
Parametric Estimation of Survival Function and Comparison Between Groups
The Likelihood Function

• The likelihood of $i$-th observation $(y_i, \delta_i, x_i)$:
  \[
  \begin{cases}
  f(y_i | X_i), & \text{if } \delta_i = 1, \text{ (not censored)} \\
  S(y_i | X_i), & \text{if } \delta_i = 0, \text{ (right – censored)}
  \end{cases}
  \]

• The likelihood of the data is
  \[
  L(\beta) \propto \prod_{i=1}^{n} \left\{ [f(y_i | X_i)]^{\delta_i} [S(y_i | X_i)]^{1-\delta_i} \right\}
  \]

  – The inference approaches based on the likelihood function can be applied.

  * See chapter by Lee on ML parameter estimation

  – Use AIC or BIC to compare nested models.

  – Graphical check of model adequacy

  * Plot $\tilde{\Lambda}(t)$ vs. $t$ for exponential models;

  * Plot $\log(\tilde{\Lambda})$ vs. $\log(t)$ for Weibull models;

  * Can also plot deviance residuals.
Accounting for Covariates: Models for Hazard Function

- General form: \( \lambda(t) = \lambda_0(t)e^{x'\beta} \)
  - \( \lambda_0(t) \): the baseline hazard
  - \( e^{x'\beta} \): multiplicative effect, independent of time

- Implications for the Survival Function:
  - \( S(t) = e^{-\Lambda(t)} = e^{-\int_0^t \lambda(s)\,ds} = e^{-\Lambda_0(t)e^{x'\beta}} \)
  - \( \log \Lambda(t) = \log[-\log S(t)] = \log \Lambda_0 + x'\beta \)
  - If the model is appropriate, a plot of \( \log[-\log S(t)_{KM}] \) for different groups yields roughly parallel lines
Accounting for Covariates: Models for Hazard Function

- Exponential distribution, no predictors:
  - \( \lambda(t) = \rho \) - constant hazard function
  - \( S(t) = e^{-\rho t} \) - survival function

- Exponential distribution, one predictor:
  - \( \lambda(t) = e^{\beta_0 + \beta_1 X} = e^{\beta_0} \cdot e^{\beta_1 X} \),
  - \( e^{\beta_0} \) plays the role of \( \lambda_0 \), i.e. the constant baseline hazard

- Weibull distribution, no predictors:
  - \( \lambda(t) = \lambda^p p t^{p-1} \) - hazard function
  - \( S(t) = e^{-(\lambda t)^p} \) - survival function

- Weibull distribution, one predictor:
  - \( \lambda(t) = p t^{p-1} \cdot e^{(\beta_0 + \beta_1 X)^p} = p t^{p-1} e^{\beta_0^p} \cdot e^{p\beta_1 X} \)
  - \( e^{\beta_0 + \beta_1 X} \) plays the role of \( \lambda \) in overall hazard
  - \( p t^{p-1} e^{\beta_0^p} \) is the baseline hazard
  - \( e^{\beta_0} \) plays the role of \( \lambda \) in the baseline hazard
More on Proportional Hazard

• Assumption of proportional hazard implies:
  
  – The hazard ratio for two subjects $i$ and $j$ with different covariates, at a given time is
    
    $$\frac{\lambda_i(t)}{\lambda_j(t)} = \frac{\lambda_0(t)}{\lambda_0(t)} \cdot \frac{e^{\beta_0 + \beta_1 X_i}}{e^{\beta_0 + \beta_1 X_j}} = e^{(X_i - X_j) \beta}$$

  – constant over time

  – only function of covariates

• Can graphically verify the assumption of proportional hazard.
  
  – Check for parallel lines of $\hat{S}^{KM}(t)$ on the complementary log-log scale

  – For the gehan dataset:

    ```R
    > plot(fit, lty=3:4, col=2:3, fun="cloglog", xlab="Remission (weeks)", ylab="log Lambda(t)"
    > legend("topleft", c("control", "6-MP"), lty=4:3, col=3:2)
    ```
Verify the Assumption of Proportional Hazard: Gehan Study

- Good agreement with the additive structure on the log-log scale
- The assumption of proportional hazard is plausible
Accounting for Covariates: Models for Survival Function

- Also called Accelerated Failure Models

\[ S(t) = S_0(t \cdot e^{X'\beta}) \]

- \( S_0 \) is the baseline survival function

- Covariates accelerate or contract the time to event

- Survival time \( T_0 = T \cdot e^{X'\beta} \) has a fixed distribution

\[
\log T_0 = \log T + X'\beta, \quad \Rightarrow \\
\log T = \log T_0 - X'\beta
\]

- Weibull (and its special case Exponential) are the only distributions that can be simultaneously (and equivalently) specified as proportional hazard and accelerated failure models.
Accelerated Failure:
Exponential

• Exponential distribution:
  \[ \lambda(t) = \rho, \quad S(t) = e^{-\rho t} \]
  \[ S_0(t) = e^{-t} - \text{survival of the standard exponential} \]

• Effect of the covariates:
  \[ S(t) = S_0(t \cdot e^{\beta_0 + \beta_1 X}), \quad e^{\beta_0 + \beta_1 X} = \rho \]
  \[ \text{If } T_0 \sim \exp(1), \text{ and } T \text{ are the observed times} \]
  \[ T_0 = T \cdot e^{\beta_0 + \beta_1 X} \]
  \[ \log T_0 = \log T + \beta_0 + \beta_1 X, \quad \text{and} \]
  \[ \log T = -\beta_0 - \beta_1 X + \log T_0 \]

• In R:
  \[ \log T = \beta_0 + \beta_1 X + \sigma \cdot \log \varepsilon, \]
  \[ \sigma \text{ is a scale parameter fixed at } \sigma = 1 \]
  \[ \varepsilon \sim \exp(1) \]
  \[ \text{use opposite sign of } \hat{\beta} \text{ to estimate survival:} \]
  \[ \hat{S}(t) = S_0(t \cdot e^{-\hat{\beta}_0 - \hat{\beta}_1 X}) \]
Accelerated Failure: Exponential

> fit.exponential <- survreg(Surv(time,cens) ~ treat, 
data=gehan, dist="exponential")

> summary(fit.exponential)

<table>
<thead>
<tr>
<th>Value</th>
<th>Std. Error</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>3.69</td>
<td>0.333</td>
<td>11.06 2.00e-28</td>
</tr>
<tr>
<td>treatcontrol</td>
<td>-1.53</td>
<td>0.398</td>
<td>-3.83 1.27e-04</td>
</tr>
</tbody>
</table>

Scale fixed at 1

Exponential distribution
Loglik(model)= -108.5 Loglik(intercept only)= -116.8
Chisq= 16.49 on 1 degrees of freedom, p= 4.9e-05

- Parameter interpretation:
  - $-\beta_{\text{treatcontrol}} = 1.53 > 0$
  - time until remission is longer for controls

- Predicted survival above $T = 10$ for trt:
  - $\exp(-10*\exp(-3.69))=0.7790189$
  - comparable to $\hat{S}_{KM}^{trt}(10) = 0.753$
Accelerated Failure: Weibull

- Weibull distribution:
  
  - $\lambda(t) = \lambda^p p t^{p-1}$ - hazard function
  
  - $S(t) = e^{-(\lambda t)^p}$ - survival function
  
  - Can be viewed as the survival function of the exponential random variable $T' = T^p \sim \text{Exp}(\lambda^p)$

- Effect of the covariates:
  
  - Identical to the exponential: $S(t') = S_0(t' \cdot e^{\beta_0+\beta_1X})$, $e^{\beta_0+\beta_1X} = \lambda^p$
  
  - If $T_0 \sim \text{exp}(1)$, and $T'$ are the observed times
    
    $T_0 = T' \cdot e^{\beta_0+\beta_1X}$
    
    $\log T_0 = \log T' + \beta_0 + \beta_1X$

  - Returning to the original notation $T' = T^p$:
    
    $\log T_0 = p \log T + \beta_0 + \beta_1X$
    
    $\log T = -\frac{1}{p} \beta_0 - \frac{1}{p} \beta_1X + \frac{1}{p} \log T_0$
Accelerated Failure: Weibull

- In R Weibull is the default distribution:
  \[ \log T = \beta_0 + \beta_1 X + \sigma \cdot \log \varepsilon, \]
  - \( \sigma \) is a scale parameter, \( \sigma = \frac{1}{p}, \varepsilon \sim \text{Exp}(1) \)
  - use \( -\hat{\beta}/\hat{\sigma} \) to estimate survival:
    \[ \hat{S}(t) = S_0 \left( t^{1/\sigma} \cdot e^{-\hat{\beta}_0/\sigma - \hat{\beta}_1/\sigma X} \right) \]

```r
> fit.weibull <- survreg(Surv(time,cens) ~ treat, data=gehan)
> summary(fit.weibull)

                 Value Std. Error  z      p
(Intercept)     3.516    0.252 13.96 2.61e-44
treatcontrol   -1.267    0.311  -4.08 4.51e-05
Log(scale)      -0.312    0.147  -2.12 3.43e-02

Scale= 0.732
Weibull distribution
Loglik(model) = -106.6  Loglik(intercept only) = -116.4
Chisq= 19.65 on 1 degrees of freedom, p= 9.3e-06
```

- Predicted survival above \( T = 10 \) for trt:
  - \( \exp(-10^{-1/0.732} \cdot \exp(-3.1516/0.732)) = 0.7308616 \)
  - comparable to \( \hat{S}_{KM}^{\text{trt}}(10) = 0.753 \)
Variable Selection and Prediction

• Compare models with and without pair as blocking factor

\[
\text{anova( survreg(Surv(time,cens) ~ treat, data=gehan),} \\
\text{survreg(Surv(time,cens)~factor(pair)+treat, data=gehan))}
\]

| Terms          | Res.Df | -2*LL  | TestDf | Deviance | P(|Chi|) |
|----------------|--------|--------|--------|----------|---------|
| treat          | 39     | 213.159| NA     | NA       | NA      |
| (pair)+treat   | 19     | 181.343| +(pair)| 20       | 31.81597| 0.04529 |

• Prediction for the median survival

  – On the linear predictor scale

> fit.weibull.nopairs <- survreg(Surv(time,cens)~treat, data=gehan)
> ?predict.survreg
> pred.contr <- predict(fit.weibull.nopairs, data.frame(treat='control'), type="uquantile", p=0.5, se=TRUE)

  – On the survival function scale

> exp( c(L=pred.contr$fit - 2*pred.contr$se.fit, U=pred.contr$fit + 2*pred.contr$se.fit) )
> L.1   U.1
5.045779 10.396147
Cox Proportional Hazards Model

A semi-parametric approach
Cox Proportional Hazards (Relative Risk) Model

• Assume:
  – $C_i$ and $T_i$ conditionally independent, given $X_i$.
  – Observe $k$ distinct exact failure times (i.e. $\delta_i = 1$); $t_1 < t_2 < \cdots < t_k$.
  – No ties in observed exact failure times (i.e., if $\delta_i = \delta_j = 1$, then $y_i \neq y_j$ for $i \neq j$).

• Define the risk set at time $t$
  
  \[ R(t) = \{ i : y_i \geq t \} = \{ \text{individuals alive right before time } t \} \]

• Cox (1975) considered the conditional probability
  – Probability of event in a small interval around $t_j$, given that the individual is in the risk set

  \[ P(\text{ind. } i \text{ failed at } [t_j, t_j + \Delta) \mid i \in R(t_j)) \approx \frac{\lambda(t_j) \Delta}{\sum_{k \in R(t_j)} \lambda(t_j) \Delta} \]
Partial Likelihood Approach to Estimate $\beta$ (Cox, 1975)

• With covariate $X_i(t)$, assume a proportional hazards model

$$\lambda(t) = \lambda_0(t) \exp X_i(t)'\beta(t)$$

where $\lambda_0(t)$ is a baseline hazard and $\exp X_i(t)'\beta(t)$ is the relative risk. Treating $\lambda_0(t)$ as a nuisance parameter, one may estimate $\beta(t)$ by maximizing the partial likelihood

$$L(\beta) = \prod_j \frac{\exp X_i(t_j)'\beta(t_j)}{\sum_{l \in R(t_j)} \exp X_l(t)'\beta(t_j)},$$

where the $i(j)$th item fails at $t_j$ and $R(t_j)$ is the risk set at $t_j$.

• Conditional on the history up to $t_j$ and the fact that one item fails at $t_j$, each term within the product is proportional to the likelihood of a multinomial model.

• $\lambda_0(t)$ is not specified, which is the non-parametric part of the model.

• $X_l(t)'\beta(t_j)$ is the parametric part of the model.
Partial Likelihood Approach to Estimate $\beta$ (Cox, 1975)

- The partial likelihood function by Cox (1975) is

$$L(\beta) = \prod_{i=1}^{n_{\text{obs}}} \frac{e^{X_i\beta}}{\sum_{k \in R(t_i)} e^{X_k\beta}} = \prod_{i=1}^{n} \left\{ \frac{e^{X_i\beta}}{\sum_{k \in R(y_i)} e^{X_k\beta}} \right\} \delta_i$$

— Cox argued that the partial likelihood has all properties of an usual likelihood, e.g., maximizing it for an optimal $\beta \implies$ maximum partial likelihood estimator (MPLE) $\hat{\beta}$

- Breslow’s Estimator of the Baseline Cumulative Hazard Rate:

$$\hat{\Lambda}_0(t) = \sum_{j: t_j \leq t} \frac{1}{\sum_{i \in R(t)} e^{X_i\hat{\beta}}} = \sum_{i: y_i \leq t} \frac{\delta_i}{\sum_{j \in R(y_i)} e^{X_j\hat{\beta}}}$$

- There are different ways to relax Assumption 3 to handle tied failure times, e.g., exact method, Efron’s method, and Breslow’s method.

<table>
<thead>
<tr>
<th>Method</th>
<th>R Options</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exact</td>
<td>method=&quot;exact&quot;</td>
<td>accurate, long</td>
</tr>
<tr>
<td>Efron’s</td>
<td>method=&quot;efron&quot;</td>
<td>approximate, better</td>
</tr>
<tr>
<td>Breslow’s</td>
<td>method=&quot;breslow&quot;</td>
<td>approximate</td>
</tr>
</tbody>
</table>


Cox Proportional Hazard

- Use exact method to account for ties

- $\hat{\beta} > 0$ as expected

```r
> fit.cox1 <- coxph(Surv(time,cens)~treat, 
method="exact", data=gehan)
> summary(fit.cox1)

n= 42, number of events= 30

coef exp(coef) se(coef)   z  Pr(>|z|)    
treatcontrol 1.6282   5.0949  0.4331 3.759 0.000170 ***

exp(coef) exp(-coef) lower .95 upper .95

treatcontrol  5.095     0.1963    2.18     11.91

Rsquare= 0.321    (max possible= 0.98 )
Likelihood ratio test= 16.25  on 1 df,  p=5.544e-05
Wald test          = 14.13  on 1 df,  p=0.0001704
Score (logrank) test = 16.79  on 1 df,  p=4.169e-05
```
Cox Proportional Hazard

• Add pair as block

```r
> fit.cox2 <- coxph(Surv(time,cens)~treat+factor(pair),
                    method="exact", data=gehan)
> summary(fit.cox2)

          coef exp(coef) se(coef)     z  Pr(>|z|)  
 treatcontrol 3.314679  27.513571 0.742620 4.463 8.06e-06 ***
 factor(pair)2 -5.015219  0.006636 1.550131-3.235  0.001215 **
 factor(pair)3 -3.598195  0.027373 1.547371-2.325  0.020053 *

Rsquare= 0.662  (max possible= 0.98 )
Likelihood ratio test= 45.51 on 21 df,  p=0.001484
Wald test = 27.42 on 21 df,  p=0.1573
Score (logrank) test = 39.73 on 21 df,  p=0.008023
```

• LR test compares log(partial likelihoods), but the test has similar properties

```r
> anova(fit.cox1, fit.cox2, test="Chisq")
Analysis of Deviance Table
   Cox model: response is Surv(time, cens)
Model 1: ~ treat
Model 2: ~ treat + factor(pair)

             loglik Chisq Df  P(>|Chi|)       
 1   -74.543                        
 2  -59.915  29.256 20  0.08283 .
```

9-34
Visualize Model Fit

- Survival curve and CI for an individual with average covariates
  - K-M curves refer to unadjusted populations, Cox curves refer to an 'average' patient
  - can make survival curves for the Cox model for specific values of covariates by providing newdata

```r
> plot(survfit(fit.cox1), lty=2:3, xlab="Remission (weeks)", ylab="Survival", main="Gehan", cex=1.5)
```
Model Diagnostics

- **Cox-Snell residuals** are most useful for examining the overall fit of a model

\[ r_{c,i} = -\log[\hat{S}(y_i|x_i)] = \hat{\Lambda}(y_i|x_i) = \hat{\Lambda}_0(y_i)e^{X_i\hat{\beta}} \]

- \( \Lambda_0(t) \) is estimated by the Breslow’s estimator.

- If the estimates \( \hat{\Lambda}_0(t) \) and \( \hat{\beta} \) were accurate, \( \{(r_{c,i}, \delta_i), i = 1, \ldots, n\} \) are right-censored observations of \( \text{Exponential}(1) \).

- For the right-censored data \( \{(r_{c,i}, \delta_i), i = 1, \ldots, n\} \), construct the Nelson-Aalen estimator \( \tilde{\Lambda}(t) \) and plot \( \tilde{\Lambda}(t) \ vs. t \).

- **Martingale residuals** can be used to determine the functional form of a covariate

\[ r_{m,i} = \delta_j - r_{c,j} \]

- Fit the Cox model with all covariates except the one of interest, and plot the martingale residuals against the covariate of interest.

- **Deviance residuals** are approximately symmetrically distributed about zero and large values may indicate outliers.

\[ r_{d,i} = \text{sign}(r_{m,i}) \sqrt{-2[r_{m,i} + \delta_i \log(r_{c,j})]} \]

9-36