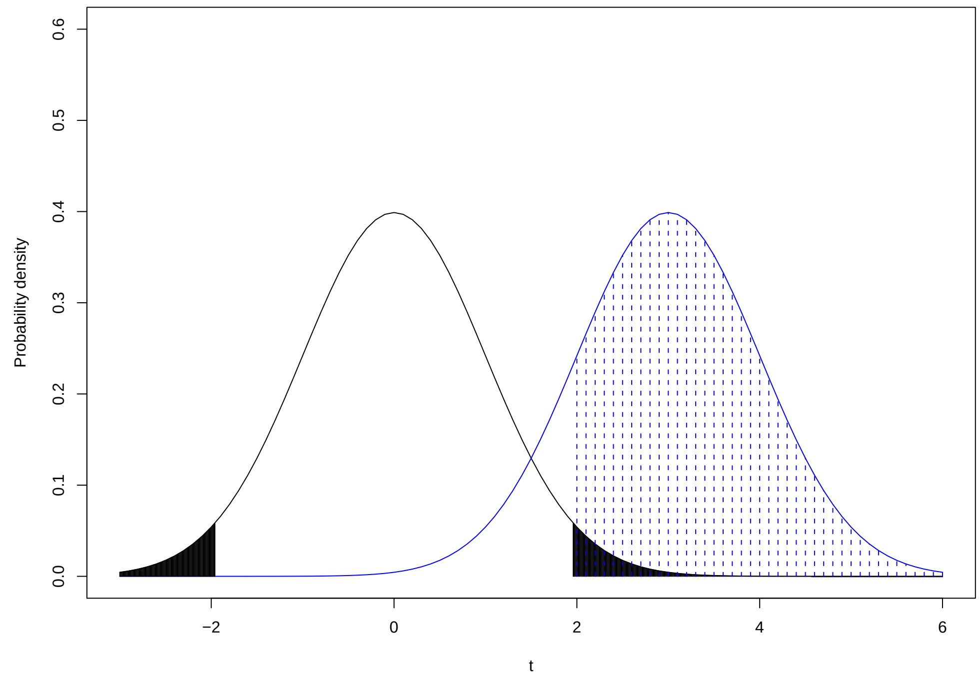
- Summary on $n^{-1/2}W$
  - Under H$_0$: $n^{-1/2}W \sim \mathcal{N}(0, A(w^2))$
  - Under H$_{1n}$: $n^{-1/2}W \sim \mathcal{N}(\xi A(\theta w), A(w^2))$
- Binary versus Time-to-event

	Binary	Time-to-event
T.S.	$\widehat{p}_1 - \widehat{p}_0$	weighted Log-rank
H$_0$	$p_1 = p_0$	$\lambda_1(t) = \lambda_0(t)$
Dist'n	$\mathcal{N}(0, \sigma_0^2)$	$\mathcal{N}(0, A(w^2))$
H$_1$	$p_1 = p_0 + d$	$\lambda_1(t) = \lambda_0(t)e^{\beta_n \theta(u)}$
Dist'n	$\mathcal{N}(d, \sigma_1^2)$	$\mathcal{N}(\xi A(\theta w), A(w^2))$

- Power



$$P = \Phi\left(\frac{\xi A(\theta w)}{A(w^2)^{1/2}} - z_{1-\alpha/2}\right)$$

- What would affect power?
  - $\frac{\xi A(\theta w)}{A(w^2)^{1/2}}$ increases, power increases
  - ξ: usually predetermined
  - $w(u)$: weight functions
- How do we choose w to maximize $\frac{A(\theta w)}{A(w^2)^{1/2}}$

- Consider $A(\cdot)$: for any constant b
  - $A[(w - b\theta)^2] \geq 0$

$$\begin{aligned} A[(w - b\theta)^2] =& A(w^2 - 2bw\theta + b^2\theta^2) \\ =& A(\theta^2)b^2 - 2bA(w\theta) + A(w^2) \geq 0 \end{aligned}$$

  - $A(w\theta)^2 - A(\theta^2)A(w^2) \leq 0$

$$\frac{A(\theta w)}{A(w^2)^{1/2}} \leq A(\theta^2)^{1/2}$$

  - equality satisfied only when

$$w(u) = \theta(u).$$

  - Cauchy-Schwarz Inequality

- Some examples of optimal $w(u)$ in nonproportional alternatives
  - Additive hazards model (Lin & Ying, 1994, BMKA):

$$\lambda(t \mid Z) = \lambda_0(t) + \beta Z \approx \lambda_0(t) e^{\frac{\beta Z}{\lambda_0(t)}} \Rightarrow w(u) = \frac{1}{\lambda_0(u)}$$

  - Accelerated hazards model (Chen & Wang, 1999, JASA):

$$\lambda(t \mid Z) = \lambda_0(te^{\beta Z}) \approx \lambda_0(t) + \lambda_0'(t) t \beta Z \Rightarrow w(u) = \frac{\lambda_0'(u) u}{\lambda_0(u)}$$

- Why $n^{1/2}\beta_n \to \xi$
  - suppose $n^k\beta_n \to \xi$ for some $k \geq 0$
  - $n^{1/2}\beta_n = n^{1/2-k}n^k\beta_n \approx n^{1/2-k}\xi$, we can verify in Term IIb
    * if $k > 1/2$, $n^{1/2}\beta_n \to 0$;
      Term II goes to 0 $\Rightarrow$ no power whatsoever
    * if $k < 1/2$, $n^{1/2}\beta_n \to \infty$;
      Term II goes to $\infty \Rightarrow$ always 100% power for any $w(u)$

3. Sample size calculation

- In practice, we have a fixed β_0 to be detected
  - $\mathrm{H}_0 : \lambda_1(t) = \lambda_0(t)$
  - $\mathrm{H}_1 : \lambda_1(t) = \lambda_0(t) e^{\beta_0 \times \theta(u)}$

- Standardized weighted Log-rank TS:
  - under H_0: $TS \sim \mathcal{N}(0, 1)$
  - under H_1:

$$TS \doteq \mathcal{N}\left(\frac{n^{1/2}\beta_0 A(\theta w)}{A(w^2)^{1/2}}, 1\right)$$

- Power $P = \text{Pr}\{|TS| \geq z_{1-\alpha/2}\} = 1 - \beta$
  - $$\frac{n^{1/2}\beta_0 A(\theta w)}{A(w^2)^{1/2}} = z_{1-\alpha/2} + z_{1-\beta} \Rightarrow n = \frac{(z_{\alpha/2} + z_\beta)^2 A(w^2)}{\beta_0 A(\theta w)^2}$$
  - $w = \theta = 1$
    * Log-rank for proportional hazards model
    * sample size

$$n = \frac{(z_{\alpha/2} + z_\beta)^2}{\beta_0^2 A(1)}$$

- what is $A(1)$?
  - recall on $A(1) = \int_0^\infty E[(Z - \mu_Z(u))^2 I(X \geq u)]\lambda_0(u)du$
  - $A(1) = \pi_Z(1 - \pi_Z)\,\mathrm{Pr}(\Delta = 1)$
- Sample size is then

$$n\,\mathrm{Pr}(\Delta = 1) = \frac{(z_{\alpha/2} + z_\beta)^2}{\beta_0^2 \pi_Z(1 - \pi_Z)}$$

- Expected # failures/events: $E_D = n\,\mathrm{Pr}(\Delta = 1)$
  - $HR = e^\beta$ is hazards ratio
  - 1-to-1 treatment-control assignment
  -

$$E_D = \frac{4(z_{\alpha/2} + z_\beta)^2}{(\log HR)^2}$$

- Example:
  - type-I error: 5%
  - power: 90%
  - $HR = 2$
  - $E_D = 42/(\log HR)^2$: 88

- Summary on comparing survival functions
  - Weighted Log-rank statistic
    - perspectives
    - asymptotics
    - power calculation
  - Alternatives
  - Yet to cover
    - stratified Log-rank
    - K-samples
    - staggered entry in sample size calculation

Chapter 5. Regression Methods

1. Proportional hazards model

2. Accelerated failure time model

3. Alternative regression models

1. Proportional hazards model

- Goal: Study the association between covariates and time-to-event outcomes

  1. covariates: time-dependent covariates
  2. outcomes: censored
  3. sampling schemes

- Breakthrough in statistics: Cox (1972, JRSS-B)

  – Peto (1972): "because he has opened up new territories to common sense"

- Semiparametric proportional hazards model

$$\lambda(t \mid Z) = \lambda_0(t) \exp(\beta^T Z)$$

  - $\lambda(t \mid Z)$: hazard function for a covariate Z
  - $Z = (Z_1, Z_2, \ldots, Z_p)^T$: p-vector covariates
  - $\lambda_0(t)$: unspecified baseline hazard function at time t; nonparametric; infinite-dimensional
  - β: p-vector regression parameters; hazards ratio; parametric; finite-dimensional

- For scalar $Z = 0/1$

$$\lambda(t \mid Z = 1) = \lambda(t \mid Z = 0)e^{\beta}$$

  - β: log of hazards ratio
  - interpretation in relative risk
    * risk: cumulative distribution/hazard function
    * $\Lambda(t \mid Z = 1) = \Lambda(t \mid Z = 0)e^{\beta}$
  - interpretation in treatment effect
    * $\beta > 0$, increased hazard implies decreased failure time
    * $\beta = 0$, no treatment effect
    * $\beta < 0$, decreased hazard implies increased failure time

- $S(t \mid Z = 1) = S(t \mid Z = 0)^{\exp(\beta)}$
  - Lehmann alternative
  - Example: $S(t \mid Z = 0) = e^{-\lambda t} \sim \mathrm{Exp}(\lambda)$

$$(e^{-\lambda t})^{e^{\beta}} = e^{-\lambda e^{\beta} t} \sim \mathrm{Exp}(\lambda e^{\beta})$$

- Critical assumption: proportionality in time-independent β

- What are the inferential problems?
  - how do we estimate β? $\lambda_0(\cdot)$?
  - what are the asymptotic properties of our estimators?
  - how do we conduct hypothesis testing regarding parameters?
  - are our estimators efficient?
  - is the model appropriate?
  - can we use the model to make prediction?
  - can we use the model to measure attribution?

- Data collected:
  - $\{(X_i, \Delta_i, Z_i), i = 1, 2, \ldots, n\}$
  - Z_i: 1-dimensional scalar
  - assumption: conditional on Z_i, T_i and C_i are independent

- Semiparametric estimation of β

- Likelihood function as if $\lambda_0(t)$ were known

$$
\begin{aligned}
\mathscr{L}(\beta) &= \prod_{i=1}^{n} f(X_i \mid Z_i)^{\Delta_i} S(X_i \mid Z_i)^{1-\Delta_i} \\
&= \prod_{i=1}^{n} \lambda(X_i \mid Z_i)^{\Delta_i} S(X_i \mid Z_i) \\
&= \prod_{i=1}^{n} \{\lambda_0(X_i) \exp(\beta Z_i)\}^{\Delta_i} \exp\left[-\int_0^{X_i} \lambda_0(u) \exp(\beta Z_i) du\right] \\
\Rightarrow l(\beta) &= \sum_{i=1}^{n} \left[\Delta_i \log \lambda_0(X_i) + \Delta_i \beta Z_i - \int_0^{X_i} \lambda_0(u) \exp(\beta Z_i) du\right]
\end{aligned}
$$

- log-likelihood:

$$l(\beta) = \sum_{i=1}^{n} \left[\Delta_i \log \lambda_0(X_i) + \Delta_i \beta Z_i - \int_0^{X_i} \lambda_0(u) \exp(\beta Z_i) du \right]$$

- Partial derivative w.r.t β

$$\begin{aligned}
\frac{\partial l}{\partial \beta} &= \sum_{i=1}^{n} \left[\Delta_i Z_i - \int_0^{X_i} \lambda_0(u) \exp(\beta Z_i) Z_i du \right] \\
&= \sum_{i=1}^{n} \left[\int_0^{\infty} Z_i dN_i(u) - \int_0^{\infty} I(X_i \geq u) Z_i \lambda_0(u) \exp(\beta Z_i) du \right] \\
&= \sum_{i=1}^{n} \int_0^{\infty} Z_i \left[dN_i(u) - Y_i(u) \lambda_0(u) \exp(\beta Z_i) du \right] \\
&= \sum_{i=1}^{n} \int_0^{\infty} Z_i dM_i(u)
\end{aligned}$$

- $E \frac{\partial l}{\partial \beta} = 0$

- $\frac{\partial l}{\partial \beta}$ is not enough because $\lambda_0(t)$ is unknown

$$\frac{\partial l}{\partial \beta} = \sum_{i=1}^{n} \int_0^{\infty} Z_i \left[dN_i(u) - Y_i(u) \lambda_0(u) \exp(\beta Z_i) du \right]$$

- Let's estimate $\lambda_0(t)$ by replacing Z_i as if β were known

$$\sum_{i=1}^{n} \int_0^{t} 1 \times \left[dN_i(u) - Y_i(u) \widehat{\lambda}_0(u) \exp(\beta Z_i) du \right] = 0$$

$$\Rightarrow \sum_{i=1}^{n} \left[dN_i(u) - Y_i(u) \widehat{\lambda}_0(u) \exp(\beta Z_i) du \right] = 0$$

$$\Rightarrow \sum_{i=1}^{n} dN_i(u) = \sum_{i=1}^{n} Y_i(u) \exp(\beta Z_i) \widehat{\lambda}_0(u) du$$

$$\Rightarrow \widehat{\lambda}_0(u) du = \frac{\sum_{i=1}^{n} dN_i(u)}{\sum_{i=1}^{n} Y_i(u) \exp(\beta Z_i)}$$

- Breslow estimator at β

$$\widehat{\Lambda}_0(t;\beta) = \int_0^t \frac{\sum_{i=1}^n dN_i(u)}{\sum_{i=1}^n Y_i(u)\exp(\beta Z_i)}$$

- cf. One-sample Nelson-Aalen estimator, i.e., $Z_i = 0$

- To solve for β, replace $\widehat{\lambda}_0(u)du$ in $\partial l/\partial\beta$:

$$\sum_{i=1}^n \int_0^\infty Z_i \left[dN_i(u) - Y_i(u)\exp(\beta Z_i) d\widehat{\Lambda}_0(u;\beta)\right]$$

- More algebra

$$
\begin{aligned}
&\sum_{i=1}^{n}\int_0^\infty Z_i\left[dN_i(u)-Y_i(u)\exp(\beta Z_i)d\widehat{\Lambda}_0(u;\beta)\right]\\
=&\sum_{i=1}^{n}\int_0^\infty Z_i\left[dN_i(u)-\frac{Y_i(u)\exp(\beta Z_i)\sum_{i=1}^{n}dN_i(u)}{\sum_{i=1}^{n}Y_i(u)\exp(\beta Z_i)}\right]\\
=&\sum_{i=1}^{n}\int_0^\infty Z_i dN_i(u)-\int_0^\infty\frac{\sum_{i=1}^{n}Z_iY_i(u)\exp(\beta Z_i)\sum_{i=1}^{n}dN_i(u)}{\sum_{i=1}^{n}Y_i(u)\exp(\beta Z_i)}
\end{aligned}
$$

- Define

$$
\bar{Z}(u;\beta)=\frac{\sum_{i=1}^{n}Z_iY_i(u)\exp(\beta Z_i)}{\sum_{i=1}^{n}Y_i(u)\exp(\beta Z_i)}\longrightarrow\frac{E[Z\exp(\beta Z)I(X\geq u)]}{E[\exp(\beta Z)I(X\geq u)]}
$$

- Then $\partial l/\partial\beta$ becomes

$$
\sum_{i=1}^{n}\int_0^\infty Z_i dN_i(u)-\int_0^\infty\bar{Z}(u;\beta)\sum_{i=1}^{n}dN_i(u)
$$

- Here is an estimating equation for β:

$$S_n(\beta) = \sum_{i=1}^{n} \int_0^{\infty} \{Z_i - \bar{Z}(u;\beta)\} \, dN_i(u)$$

  - under the proportional hazards model, this is the partial score equation
  - in general, this is called quasi partial score equation
  - this way of getting appropriate estimating functions is called Quasi Partial Scoring (QPS).
  - Estimator $\widehat{\beta}_n$: $S_n(\widehat{\beta}_n) = 0$

- Primary inferential questions: for true value $\beta = \beta_0$
  - consistent: $\widehat{\beta}_n \to_P \beta_0$
  - asymptotically normal: $\sqrt{n}(\widehat{\beta}_n - \beta_0) \to_D \mathcal{N}(0, \sigma^2)$
- Secondary inferential questions
  - optimality
  - asymptotic properties for Breslow estimator

- From estimating function to estimators:

$$S_n(\widehat{\beta}_n) - S_n(\beta_0) = -S_n(\beta_0) \approx S'(\beta_0)(\widehat{\beta}_n - \beta_0)$$

  – $S_n(\beta_0)$: for zero mean and variation

  – $S'(\beta_0)$: for efficiency

- Recall: $\bar{Z}(u;\beta) = \frac{\sum_{i=1}^n Z_i Y_i(u)\exp(\beta Z_i)}{\sum_{i=1}^n Y_i(u)\exp(\beta Z_i)}$

- What is $S_n(t;\beta_0)$?

$$S_n(t;\beta_0) = \sum_{i=1}^n \int_0^t \{Z_i - \bar{Z}(u;\beta_0)\}\, dN_i(u)$$

$$= \sum_{i=1}^n \int_0^t \{Z_i - \bar{Z}(u;\beta_0)\} [dN_i(u) - Y_i(u)\lambda_0(u)\exp(\beta_0 Z_i)du$$

$$+ \sum_{i=1}^n \int_0^t \{Z_i - \bar{Z}(u;\beta_0)\}\, Y_i(u)\exp(\beta_0 Z_i)\lambda_0(u)du$$

$$= \sum_{i=1}^n \int_0^t \{Z_i - \bar{Z}(u;\beta_0)\}\, dM_i(u;\beta_0)$$

$$+ \int_0^t \left\{\sum_{i=1}^n Z_i Y_i(u)\exp(\beta_0 Z_i) - \sum_{i=1}^n Y_i(u)\exp(\beta_0 Z_i)\bar{Z}(u;\beta_0)\right\} \lambda_0(u)du$$

$$= \sum_{i=1}^n \int_0^t \{Z_i - \bar{Z}(u;\beta_0)\}\, dM_i(u;\beta_0)$$

- Predictable variation of $U_n(t) = n^{-1/2} S_n(t; \beta_0)$

$$< U_n, U_n > (t) = n^{-1} \sum_{i=1}^{n} \int_0^t \{Z_i - \bar{Z}(u)\}^2 Y_i(u) \exp(\beta_0 Z_i) \lambda_0(u) du$$

$$\to \int_0^t E\left[\{Z - \mu_Z(u)\}^2 Y(u) \exp(\beta_0 Z)\right] \lambda_0(u) du$$

- Properties of $S_n(t; \beta_0)$:

1. $ES_n(t; \beta_0) = 0$

2. independent increment for $s \leq t$

$$\text{cov}[S_n(s; \beta_0)\{S_n(t; \beta_0) - S_n(t; \beta_0)\}] = 0$$

3. $\text{var}\{n^{-1/2} S_n(t; \beta_0)\} = E < U_n, U_n > (t)$

- Consistency (sketch of a proof)
  - roughly, $-S_n(\beta_0) \approx S'(\beta_0)(\widehat{\beta}_n - \beta_0)$
  - by WLLN, $n^{-1} S_n(\beta_0) \to_P 0$
  - for arbitrary $\beta \neq \beta_0$

    $$
    \begin{aligned}
    n^{-1} S_n(\beta) - n^{-1} S_n(\beta_0) \\
    &= - n^{-1} \sum_{i=1}^{n} \int_0^t \{\bar{Z}(u;\beta) - \bar{Z}(u;\beta_0)\}\, dN_i(u) \\
    &\to_P \Gamma(\beta) = - \int_0^t \{\mu(u;\beta) - \mu(u;\beta_0)\}\, d\,\mathrm{Pr}\{X \le u, \Delta = 1\}
    \end{aligned}
    $$

  - we can verify

    $$
    \frac{\partial \mu(u;\beta_0)}{\partial \beta} = \frac{E[\{Z - \mu(u,\beta_0)\}^2 \exp(\beta_0 Z) Y(u))]}{E[\exp(\beta_0 Z) Y(u))]} > 0
    $$

    provided Z is not constant

- then for $\Gamma(\beta)$
  1. $n^{-1}S_n(\beta) \to_P \Gamma(\beta)$
  2. $\Gamma(\beta)$ and $n^{-1}S_n(\beta)$ are strictly decreasing
  3. $\Gamma(\beta_0) = 0$
- similar to the Glivenko-Cantelli Lemma $\Rightarrow n^{-1}S_n(\beta) \to_P \Gamma(\beta)$ uniformly in $\beta \in U(\beta_0)$
- for any $\epsilon > 0$, there exists sufficiently large n, s.t.

$$S_n(\beta_0 - \epsilon) > 0, S_n(\beta_0 + \epsilon) < 0$$

  therefore, $\widehat{\beta}_n \in (\beta_0 - \epsilon, \beta_0 + \epsilon)$ with sufficiently large probability
- $\widehat{\beta}_n \to_P \beta_0$