Survival Analysis: Time To Event Modelling

Sudipta Das

Assistant Professor,
Department of Computer Science,
Ramakrishna Mission Vivekananda Educational & Research Institute

Outline I

Introduction I

- By Modeling the hazard rate, we can understand how quickly individuals of a certain age are experiencing the event of interest.
- Therefore, we try to model the hazard function based on covariates/treatment/confounder.
- The major approach to modeling the effects of covariates on survival is to model the conditional hazard rate as a function of the covariates.
- Two general classes of models have been used to relate covariate effects to survival,
  - the family of multiplicative hazard models and
  - the family of additive hazard rate models.

Introduction II

- Let X denote the time to some event.
- Data, based on a sample of size n,

$$\left(T_j, \delta_j, \underline{Z}_j(t)\right), \; j = 1, \ldots, n$$

where

- T_j is the time on study for the jth patient,
- δ_j is the event indicator for the jth patient and
- $\underline{Z}_j(t) = (Z_{j1}(t), \ldots, Z_{jp}(t))^T$ is the vector of covariates or risk factors for the jth individual at time t

Introduction III

- $Z_{jk}(t)$'s, $k = 1, \ldots, p$, may affect the survival distribution of X.
- $Z_{jk}(t)$'s, $k = 1, \ldots, p$, may be
  - time-dependent covariates, such as
    - current disease status, serial blood pressure measurements, etc.,
  - constant values known at time 0, such as
    - sex, treatment group, race, initial disease state, etc.
- We shall consider the fixed-covariate case where

$$\underline{Z}_j(t) = \underline{Z}_j = (Z_{j1}, \ldots, Z_{jp})^T.$$

Introduction IV

- Family of multiplicative hazard rate models
  - the conditional hazard rate of an individual with covariate vector $\underline{z}$ is a product of a baseline hazard rate $h_0(t)$ and a non-negative function of the covariates, $c(\underline{\beta}^T \underline{z})$, that is,

$$h(t|\underline{z}) = h_0(t)c(\underline{\beta}^T \underline{z}),$$

    where $\underline{\beta} = [\beta_1, \ldots, \beta_p]^T$ is a parameter vector.
  - $h_0(t)$ may have a specified parametric form or it may be left as an arbitrary non-negative function.
  - $c(\cdot)$ can be any non-negative link function

Introduction V

- This is called a semi-parametric model
  - when a parametric form is assumed only for the covariate effect and
  - the baseline hazard rate is treated non-parametrically.

- Survival function:

$$
\begin{aligned}
S(t|\underline{z}) &= e^{-\int_0^t h(u|\underline{z})du} \\
&= e^{-\int_0^t h_0(u)c(\underline{\beta}^T\underline{z})du} \\
&= \left[e^{-\int_0^t h_0(u)du}\right]^{c(\underline{\beta}^T\underline{z})} \\
&= [S_0(t)]^{c(\underline{\beta}^T\underline{z})}
\end{aligned}
$$

  - $S_0(t)$ is called baseline survival function

Introduction VII

- A common link function uses in most applications is

$$
\begin{aligned}
c(\underline{\beta}^T \underline{z}) &= e^{\underline{\beta}^T \underline{z}} \\
&= e^{\sum_{k=1}^{p} \beta_k z_k}.
\end{aligned}
$$

  - Note that $e^{\sum_{k=1}^{p} \beta_k z_k}$ is always positive.
  - Cox (1972) proportional hazards model.

Proportional Hazards Model I

- Cox's Regression Model

$$h(t|\underline{z}) = h_0(t)e^{\underline{\beta}^T\underline{z}}.$$

$$\underbrace{\text{log of hazard for given covariate profile}}_{\log h(t|\underline{z})}$$
$$= \underbrace{\text{log of baseline hazard}}_{\log h_0(t)} + \underbrace{\text{linear combination of covariates}}_{\underline{\beta}^T\underline{z}}$$

- $\underline{\beta}$ describe the rate of change of log-hazard with covariates.

Proportional Hazards Model II

- Survival function

$$
\begin{aligned}
S(t|z) &= [S_0(t)]^{e^{\beta^T z}} \\
&= [S_0(t)]^{e^{\sum_{k=1}^{p} \beta_k z_k}},
\end{aligned}
$$

where

$$S_0(t) = e^{-\int_0^t h_0(u)du} = e^{-H_0(t)}.$$

Proportional Hazards Model III

- The Cox model is often called a proportional hazards model.
- For two individuals with covariate values Z and Z^* , the ratio of their hazard rates is constant or independent of time.

$$\begin{aligned} \frac{h(t|Z)}{h(t|Z^*)} &= \frac{h_0(t)e^{\sum_{k=1}^{p}\beta_k z_k}}{h_0(t)e^{\sum_{k=1}^{p}\beta_k z_k^*}} \\ &= \exp\left[\sum_{k=1}^{p}\beta_k(z_k - z_k^*)\right] \end{aligned}$$

  - Hazard rates are proportional

Proportional Hazards Model IV

- This ratio is called the relative risk (hazard ratio) of an individual with risk factor Z having the event as compared to an individual with risk factor Z^*.
- In particular, keeping all other covariates have the same value, if Z_1 indicates the treatment effect
  - ($Z_1 = 1$ if treatment and $Z_1 = 0$ if placebo)
- then,

$$h(t|Z)/h(t|Z^*) = e^{\beta_1},$$

is *the risk of having the event if the individual received the treatment* relative to *the risk of having the event should the individual have received the placebo*.

Partial Likelihoods for Distinct-Event Time Data: Construction I

- Data:

$$(T_j, \delta_j, \underline{Z}_j),\ j = 1(1)n$$

- Ordered event times:

$$t_1 < t_2 < \ldots < t_D$$

- kth covariate of the individual whose failure time is t_i:

$$Z_{(i)k}$$

- Risk set at t_i :

$$R(t_i)$$

  - the set of all individuals who are still under study at a time just prior to t_i

Partial Likelihoods for Distinct-Event Time Data: Construction II

- Likelihood contribution of the individual whose failure time is t_i

$$
\begin{aligned}
L_i(\underline{\beta}) &= P\,[\text{ individual dies at } t_i \mid \text{one death at } t_i] \\
&= \frac{P\,[\text{ individual dies at } t_i \mid \text{survival to } t_i]}{P\,[\text{ one death at } t_i \mid \text{survival to } t_i]} \\
&= \frac{h[t_i|\underline{Z}_{(i)}]}{\sum_{j\in R(t_i)} h[t_i|\underline{Z}_j]} \\
&= \frac{h_0[t_i]e^{\underline{\beta}^T\underline{Z}_{(i)}}}{\sum_{j\in R(t_i)} h_0[t_i]e^{\underline{\beta}^T\underline{Z}_j}} = \frac{e^{\underline{\beta}^T\underline{Z}_{(i)}}}{\sum_{j\in R(t_i)} e^{\underline{\beta}^T\underline{Z}_j}}
\end{aligned}
$$

Partial Likelihoods for Distinct-Event Time Data: Construction III

- The Cox partial likelihood over all deaths

$$L(\underline{\beta}) = \prod_{i=1}^{D} L_i(\underline{\beta}) = \prod_{i=1}^{D} \frac{e^{\underline{\beta}^T \underline{Z}_{(i)}}}{\sum_{j \in R(t_i)} e^{\underline{\beta}^T \underline{Z}_j}}$$

Partial Likelihoods for Distinct-Event Time Data: Construction IV

- The Cox partial likelihood can *also* be derived as a profile likelihood from the full censored-data likelihood.
- Derivation
  - The complete censored-data likelihood

$$\begin{aligned} L[\underline{\beta}, h_0(t)] &= \prod_{j=1}^{n} \left\{ \left[h(T_j|\underline{Z}_j)\right]^{\delta_j} S(T_j|\underline{Z}_j) \right\} \\ &= \prod_{j=1}^{n} \left\{ \left[h_0(T_j)e^{\underline{\beta}^T \underline{Z}_j}\right]^{\delta_j} e^{-H_0(T_j)e^{\underline{\beta}^T \underline{Z}_j}} \right\} \end{aligned}$$

  - Now, for a fixed $\underline{\beta}$, the profile likelihood of the estimator $h_0(t)$

$$L_{\underline{\beta}}[h_0(t)] = \left[\prod_{i=1}^{D} h_0(t_i)e^{\underline{\beta}^T \underline{Z}_{(i)}}\right] e^{-\left[\sum_{j=1}^{n} H_0(T_j)e^{\underline{\beta}^T \underline{Z}_j}\right]}$$

Partial Likelihoods for Distinct-Event Time Data: Construction V

- Note that, this function is maximal when $h_0(t) = 0$ except for times at which the events occurs.
- Let
$$h_{0i} = h_0(t_i), \; i = 1, \ldots, D$$
- So
$$H_0(T_j) = \sum_{t_i \leq T_j} h_{0i}.$$

Partial Likelihoods for Distinct-Event Time Data: Construction VI

- Thus,

$$
\begin{aligned}
L_{\underline{\beta}}[h_{01},\ldots,h_{0D}] &= \left[\prod_{i=1}^{D} h_0(t_i) e^{\underline{\beta}^T \underline{Z}_{(i)}}\right] e^{-\left[\sum_{j=1}^{n}\sum_{i:t_i\leq T_j} h_{0i} e^{\underline{\beta}^T \underline{z}_j}\right]} \\
&= \prod_{i=1}^{D}\left\{ h_{0i} e^{\underline{\beta}^T \underline{Z}_{(i)}} \times e^{-h_{0i}\left[\sum_{j\in R(t_i)} e^{\underline{\beta}^T \underline{z}_j}\right]}\right\} \\
&= \prod_{i=1}^{D}\left\{ h_{0i} e^{-h_{0i}\left[\sum_{j\in R(t_i)} e^{\underline{\beta}^T \underline{z}_j}\right]} \times e^{\underline{\beta}^T \underline{Z}_{(i)}}\right\} \\
&\propto \prod_{i=1}^{D}\left\{ h_{0i} e^{-h_{0i}\left[\sum_{j\in R(t_i)} e^{\underline{\beta}^T \underline{z}_j}\right]}\right\}
\end{aligned}
$$

Partial Likelihoods for Distinct-Event Time Data: Construction VII

- Therefore, the profile maximum likelihood estimator of h_{0i} is

$$\hat{h}_{0i} = \frac{1}{\sum_{j \in R(t_i)} e^{\underline{\beta}^T \underline{Z}_j}}$$

- Also, the estimate of $H_0(t)$ is

$$\hat{H}_0(t) = \sum_{t \leq t_i} \frac{1}{\sum_{j \in R(t_i)} e^{\underline{\beta}^T \underline{Z}_j}}$$

  - This is called Breslow's estimator of the baseline cumulative hazard rate in the case of, at most, one death at any time

Partial Likelihoods for Distinct-Event Time Data: Construction VIII

- Substituting $\hat{H}_0(t)$ in complete censor data likelihood, we get the profile likelihood proportional to the partial likelihood of $\underline{\beta}$ as

$$
\begin{aligned}
L(\underline{\beta}) &= \prod_{i=1}^{D} \left\{ \frac{e^{\underline{\beta}^T \underline{Z}_{(i)}}}{\sum_{j \in R(t_i)} e^{\underline{\beta}^T \underline{Z}_j}} e^{-1} \right\} \\
&\propto \prod_{i=1}^{D} \frac{\exp\left[\sum_{k=1}^{p} \beta_k Z_{(i)k}\right]}{\sum_{j \in R(t_i)} \exp\left[\sum_{k=1}^{p} \beta_k Z_{jk}\right]}
\end{aligned}
$$

- Note that
  - the numerator of the likelihood depends only on information from the individual who experiences the event,
  - the denominator utilizes information about all individuals who have not yet experienced the event (including some individuals who will be censored later).

Partial Likelihoods for Distinct-Event Time Data: Estimation I

- The (partial) log-likelihood

$$
\begin{aligned}
l(\underline{\beta}) &= \log L(\underline{\beta}) \\
&= \sum_{i=1}^{D} \sum_{k=1}^{p} \beta_k Z_{(i)k} - \sum_{i=1}^{D} \log \left(\sum_{j \in R(t_i)} \exp \left[\sum_{k=1}^{p} \beta_k Z_{jk} \right] \right)
\end{aligned}
$$

Partial Likelihoods for Distinct-Event Time Data: Estimation II

- Thus, the *score functions* are

$$
\begin{aligned}
U_h(\underline{\beta}) &= \frac{\delta}{\delta\beta_h} l(\underline{\beta}), \; h = 1, \ldots, p \\
&= \sum_{i=1}^{D} Z_{(i)h} - \sum_{i=1}^{D} \frac{\sum_{j\in R(t_i)} Z_{jh} \exp\left[\sum_{k=1}^{p} \beta_k Z_{jk}\right]}{\sum_{j\in R(t_i)} \exp\left[\sum_{k=1}^{p} \beta_k Z_{jk}\right]}
\end{aligned}
$$

  - Partial derivatives of log-likelihood with respect to the parameters
  - Note that: -
    - $E[U_h(\beta)] = E\left[\frac{\delta}{\delta\beta_h} l(\underline{\beta})\right] = 0$ for all $h = 1, \ldots, p$
    - $Cov[U_g(\beta)U_h(\beta)] = E\left[\frac{\delta}{\delta\beta_g} l(\underline{\beta}) \frac{\delta}{\delta\beta_h} l(\underline{\beta})\right] = -E\left[\frac{\delta^2}{\delta\beta_g\delta\beta_h} l(\underline{\beta})\right]$

Partial Likelihoods for Distinct-Event Time Data: Estimation III

- The information matrix is $\mathcal{I}(\underline{\beta}) = \left[\mathcal{I}_{gh}(\underline{\beta})\right]_{p\times p}$, where the $(g,h)^{th}$ element is

$$\begin{aligned}
\mathcal{I}_{gh}(\underline{\beta}) &= -\frac{\delta^2}{\delta\beta_g\delta\beta_h} l(\underline{\beta}) = -\frac{\delta}{\delta\beta_g} U_h(\underline{\beta}) \\
&= \sum_{i=1}^{D} \frac{\sum_{j\in R(t_i)} Z_{jg}Z_{jh} \exp\left[\sum_{k=1}^{p} \beta_k Z_{jk}\right]}{\sum_{j\in R(t_i)} \exp\left[\sum_{k=1}^{p} \beta_k Z_{jk}\right]} \\
&- \sum_{i=1}^{D} \frac{\sum\limits_{j\in R(t_i)} Z_{jg} \exp\left[\sum\limits_{k=1}^{p} \beta_k Z_{jk}\right]}{\sum\limits_{j\in R(t_i)} \exp\left[\sum\limits_{k=1}^{p} \beta_k Z_{jk}\right]} \times \frac{\sum\limits_{j\in R(t_i)} Z_{jh} \exp\left[\sum\limits_{k=1}^{p} \beta_k Z_{jk}\right]}{\sum\limits_{j\in R(t_i)} \exp\left[\sum\limits_{k=1}^{p} \beta_k Z_{jk}\right]}
\end{aligned}$$

  - Negative of the matrix of second derivatives of the log likelihood

Partial Likelihoods for Distinct-Event Time Data: Estimation IV

- The (partial) maximum likelihood estimates $\hat{\underline{\beta}} = \underline{b}$ are found by solving the set of p nonlinear equations

$$U_h(\underline{\beta}) = 0, \;\; h = 1, \ldots, p.$$

- The estimated standard error of the estimates, i.e., $\hat{se}(\underline{b})$ can be found from the inverse of the information matrix calculated at $\underline{\beta} = \underline{b}$, i.e.,

$$\mathcal{I}^{-1}(\underline{\beta})|_{\underline{\beta}=\underline{b}}$$

  - Note that: - *mle* is an efficient estimator for large sample

Partial Likelihoods for Distinct-Event Time Data: Testing I

- There are three main tests for hypotheses about regression parameters $\underset{\sim}{\beta}$
  - Wald's test
  - The likelihood ratio test
  - The scores test
- General setup
  - Let $\underset{\sim}{b} = (b_1, \ldots, b_p)^T$ denote the (partial) maximum likelihood estimates of $\underset{\sim}{\beta}$ and
  - let $\mathcal{I}(\underset{\sim}{\beta})$ be the $p \times p$ information matrix evaluated at $\underset{\sim}{\beta}$.

Partial Likelihoods for Distinct-Event Time Data: Testing II

- Wald's test
  - It is based on the result that, for large samples, $\underline{b}$ has a $p-$variate normal distribution with mean $\underline{\beta}$ and variance-covariance estimated by $\mathcal{I}^{-1}(b)$, i.e.
$$\underline{b} \sim N_p\left(\underline{\beta}, \mathcal{I}^{-1}(\underline{b})\right).$$

Partial Likelihoods for Distinct-Event Time Data: Testing III

- Wald's test (Contd.)
  - Null Hypothesis,

$$H_0 : \underline{\beta} = \underline{\beta}_0$$

  - Test statistics,

$$X_W^2 = (\underline{b} - \underline{\beta}_0)^T \mathcal{I}(\underline{b})(\underline{b} - \underline{\beta}_0)$$

  - Under H_0

$$X_W^2 \sim \chi^2(p), \text{ for large } n.$$

Partial Likelihoods for Distinct-Event Time Data: Testing IV

- The likelihood ratio test
  - Null Hypothesis,

$$H_0 : \underline{\beta} = \underline{\beta}_0$$

  - Test statistics,

$$\begin{aligned} X^2_{LR} &= -2 \ln \left(\frac{\max\limits_{\beta \in \{\beta_0\}} L(\beta)}{\max\limits_{\beta \in \mathcal{R}^p} L(\beta)} \right) \\ &= 2 \left[l(\underline{b}) - l(\underline{\beta}_0) \right] \end{aligned}$$

,

  - Under H_0

$$X^2_{LR} \sim \chi^2(p), \text{ for large } n.$$

Partial Likelihoods for Distinct-Event Time Data: Testing V

- The scores test
  - It is based on the result that, for large samples,

$$U(\underline{\beta}) = \left[U_1(\underline{\beta}), \ldots, U_p(\underline{\beta})\right]^T$$

is asymptotically $p-$variate normal with mean 0 and covariance $\mathcal{I}(\underline{\beta})$, i.e.,

$$U(\underline{\beta}) \sim N_p\left(\underline{0}, \mathcal{I}(\underline{\beta})\right)$$

Partial Likelihoods for Distinct-Event Time Data: Testing VI

- The scores test (Contd.)
  - Null Hypothesis,

$$H_0 : \underline{\beta} = \underline{\beta}_0$$

  - Test statistics,

$$X^2_{SC} = \left[U(\underline{\beta}_0)\right]^T \mathcal{I}^{-1}(\underline{\beta}_0)\left[U(\underline{\beta}_0)\right]$$

  - Under H_0

$$X^2_{SC} \sim \chi^2(p), \text{ for large } n$$

- See *Example 8.1* of Page 255.

Partial Likelihoods for Event Time Data with Ties I

- Data: $(T_j, \delta_j, \underline{Z}_j),\ j = 1(1)n$
- Distinct ordered event times:

$$t_1 < t_2 < \ldots < t_D$$

  - let the number of deaths at t_i be d_i
  - let the set of all individuals who die at time t_i be $\mathcal{D}_i$
  - let the sum of the vectors $\underline{Z}_j$ over all individuals who die at t_i be $\underline{s}_i$,

$$\underline{s}_i = \sum_{j \in \mathcal{D}_i} \underline{Z}_j$$

- Risk set at t_i : R_i
  - the set of all individuals at risk just prior to t_i

Partial Likelihoods for Event Time Data with Ties II

- There are several suggestions for constructing the partial likelihood when there are ties among the event times.
  - Breslow's Likelihood
  - Efron's Likelihood
  - Discrete Likelihood
- When there are no ties between the event times, all the three likelihoods reduce to the partial likelihood in the previous section.

Partial Likelihoods for Event Time Data with Ties III

- Breslow's Likelihood:

$$
\begin{aligned}
L_1(\underline{\beta}) &= \prod_{i=1}^{D} \prod_{j=1}^{d_i} \frac{e^{\underline{\beta}^T \underline{Z}_j}}{\sum_{k \in R_i} e^{\underline{\beta}^T \underline{Z}_k}} \\
&= \prod_{i=1}^{D} \frac{e^{\underline{\beta}^T \underline{s}_i}}{\left[\sum_{k \in R_i} e^{\underline{\beta}^T \underline{Z}_k}\right]^{d_i}}
\end{aligned}
$$

- Note:
  - Breslow's likelihood considers each of the d_i events at a given time as distinct,
  - Thus it constructs their individual contribution to the likelihood function, and obtains the overall likelihood by multiplying these contributions over all events at time t_i.
  - When there are few ties, this approximation works quite well.

Partial Likelihoods for Event Time Data with Ties IV

- Efron's Likelihood

$$L_2(\underline{\beta}) = \prod_{i=1}^{D}\prod_{j=1}^{d_i} \frac{e^{\underline{\beta}^T \underline{Z}_j}}{\left[\sum_{k\in R_i} e^{\underline{\beta}^T \underline{Z}_k} - \frac{j-1}{d_i}\sum_{k\in\mathcal{D}_i} e^{\underline{\beta}^T \underline{Z}_k}\right]}$$

$$= \prod_{i=1}^{D} \frac{e^{\underline{\beta}^T \underline{s}_i}}{\prod_{j=1}^{d_i}\left[\sum_{k\in R_i} e^{\underline{\beta}^T \underline{Z}_k} - \frac{j-1}{d_i}\sum_{k\in\mathcal{D}_i} e^{\underline{\beta}^T \underline{Z}_k}\right]}$$

- Note:
  - Efron's likelihood is closer to the correct partial likelihood based on a discrete hazard model than Breslow's likelihood.
  - When the number of ties is small, Efron's and Breslow's likelihoods are quite close.

- Discrete Likelihood

$$L_3(\underline{\beta}) = \prod_{i=1}^{D} \frac{e^{\underline{\beta}^T \underline{s}_i}}{\left[\sum_{q \in Q_i} e^{\underline{\beta}^T \underline{s}_q^*}\right]}$$

  - Q_i denote the set of all subsets of d_i individuals who could be selected from the risk set R_i.
    - Each element of Q_i is a d_i-tuple of individuals who could have been one of the d_i failures at time t_i.
  - $q = (q_1, \ldots, q_{d_i}) \in Q_i$ and $\underline{s}_q^* = \sum_{j=1}^{d_j} \underline{Z}_{qj}$.

- *Example 8.4*: A study to assess the time to first exit-site infection (in months) in patients with renal insufficiency was conducted. 43 patients utilized a surgically placed catheter and 76 patients utilized a percutaneous placement of their catheter. Catheter failure was the primary reason for censoring.
  To apply a proportional hazards regression, let $Z = 1$ if the patient has a percutaneous placement of the catheter, and 0 otherwise.
  There are 6 deaths at time 0.5. All 6 deaths have $Z = 1$, and there are 76 patients at risk with $Z = 1$ and 43 patients at risk with $Z = 0$
- Likelihood contribution at $t_1 = 0.5$,
  - Berslow: $\dfrac{e^{6\beta}}{[43 + 76e^{\beta}]^6}$
  - Efron: $\dfrac{e^{6\beta}}{\prod_{j=1}^{6}\left[43 + 76e^{\beta} - \frac{j-1}{6}(6e^{\beta})\right]}$
  - Discrete:

$$\frac{e^{6\beta}}{\left[\binom{43}{6}+\binom{43}{5}\binom{76}{1}e^{\beta}+\binom{43}{4}\binom{76}{2}e^{2\beta}+\binom{43}{3}\binom{76}{3}e^{3\beta}+\binom{43}{2}\binom{76}{4}e^{4\beta}+\binom{43}{1}\binom{76}{5}e^{5\beta}+\binom{76}{6}e^{6\beta}\right]}$$

Estimation of the Survival Function based on Breslow's estimator I

- To construct this estimator, at first, fit a proportional hazards model to the data
  - and obtain the partial maximum likelihood estimators $\underline{b}$
  - and the estimated covariance matrix $\hat{V}(\underline{b})$ from the inverse of the information matrix.
- Let $t_1 < t_2 < \cdots < t_D$ denote the distinct death times and
- let d_i be the number of deaths at time t_i.
- Let

$$W(t_i, \underline{b}) = \sum_{j \in R(t_i)} e^{\sum_{h=1}^{p} b_h Z_{jh}}$$

Estimation of the Survival Function based on Breslow's estimator II

- Thus, the estimator of the cumulative baseline hazard rate $H_0(t)$ is

$$\hat{H}_0(t) = \sum_{t_i \le t} \frac{d_i}{W(t_i, \underline{b})}$$

  - It is a step function with jumps at the observed death times.
  - This estimator reduces to the Nelson-Aalen estimator, when there are no covariates present,

- The estimator of the baseline survival function, $S_0(t) = e^{-H_0(t)}$ is

$$\hat{S}_0(t) = e^{-\hat{H}_0(t)}$$

  - This is an estimator of the survival function of an individual with a baseline set of covariate values, $\underline{Z} = 0$

Estimation of the Survival Function based on Breslow's estimator III

- To estimate the survival function for an individual with a covariate vector $\underline{Z} = \underline{Z}_0$, we use the estimator

$$\hat{S}(t|\underline{Z} = \underline{Z}_0) = \left[\hat{S}_0(t)\right]^{\exp(b^T \underline{Z}_0)}.$$

- Under mild regularity conditions the estimator $\hat{S}(t|\underline{Z} = \underline{Z}_0)$, for fixed t, has an asymptotic normal distribution with mean

$$S(t|\underline{Z} = \underline{Z}_0).$$

Estimation of the Survival Function based on Breslow's estimator IV

- The variance of the asymptotic normal distribution can be estimated by

$$\hat{V}\left[\hat{S}(t|\underline{Z}=\underline{Z}_0)\right]=\left[\hat{S}(t|\underline{Z}=\underline{Z}_0)\right]^2\left[Q_1(t)+Q_2(t;\underline{Z}_0)\right],$$

where

- $Q_1(t)=\sum_{t_i\leq t}\frac{d_i}{W(t_i,\underline{b})^2}$ and
- $Q_2(t;\underline{Z}_0)=\left[\underline{Q}_3(t;\underline{Z}_0)\right]^T\hat{V}(\underline{b})\left[\underline{Q}_3(t;\underline{Z}_0)\right]$ with
- $\underline{Q}_3(t;\underline{Z}_0)=\left[Q_3(t;\underline{Z}_0)_1,\ldots,Q_3(t;\underline{Z}_0)_k,\ldots,Q_3(t;\underline{Z}_0)_p\right]^T$ where
  - $Q_3(t;\underline{Z}_0)_k=\sum_{t_i\leq t}\left(\frac{W^{(k)}(t_i;\underline{b})}{W(t_i;\underline{b})}-Z_{0k}\right)\left(\frac{d_i}{W(t_i;\underline{b})}\right)$ and
  - $W^{(k)}(t_i;\underline{b})=\sum_{j\in R(t_i)}Z_{jk}e^{\underline{b}^T\underline{Z}_j}$

Estimation of the Survival Function based on Breslow's estimator V

- Note:
  - Q_1 is an estimator of the variance of $\hat{H}_0(t)$ if $\underline{b}$ were the true value of $\underline{\beta}$.
    - Recall NA estimates:

$$\tilde{H} = \sum_{t_i \le t} \frac{d_i}{Y_i} \text{ and } \hat{\sigma}^2_{\tilde{H}} = \sum_{t_i \le t} \frac{d_i}{Y_i^2}$$

  - Q_2 reflects the uncertainty in the estimation process added by estimating $\underline{\beta}$.
  - $Q_3(t, \underline{Z}_0)$ is large when $\underline{Z}_0$ is far from the average covariate in the risk set.

Estimation of the Survival Function based on Breslow's estimator VI

- Using this variance estimate, point-wise confidence intervals for the survival function can be constructed for $S(t|\underline{Z} = \underline{Z}_0)$ using the techniques discussed earlier
- As we have seen earlier, the log-transformed or arcsine-square-root-transformed intervals perform better than the naive, linear, confidence interval.

Example I

- Example based on bank credit data
- Cox's Proportional Hazard model
  - Using 5 covariates; (Age, Amount, InstallmentRatePercentage, NumberExistingCredits and NumberPeopleMaintenance)

$$\hat{\beta} = [-9.72\times10^{-3}, -1.96\times10^{-4}, -9.09\times10^{-2}, 2.54\times10^{-2}, -1.06\times10^{-2}]^T$$

  - Using 2 covariates; (Amount and InstallmentRatePercentage)

$$\hat{\beta} = [-1.99\times10^{-4}, -1.10\times10^{-1}]^T$$

  - Baseline Cumulative Hazards: **FIGURE 8a**

Example II

- Comparing predictions for
  - *Amount* at
    - mean + sd ($X = \mu + \sigma$) and
    - mean - sd ($X = \mu - \sigma$)
  - *InstallmentRatePercentage* is kept constant at mean
- **FIGURE 8b**