Predicted and Conditional Marginals for Cox's Proportional Hazards Model Using SUDAAN®

Elvessa D. Aragon-Logan, G. Gordon Brown, Babubhai Shah, Beth Barnwell

Abstract

Obtaining adjusted treatment means for regression models is often relevant to the researcher. The primary goal of our work is to produce functional forms for two types of adjusted treatment means, predicted marginals and conditional marginals, for Cox's proportional hazards model for cluster-correlated data with tied event times. We present two methods, Breslow's and Efron's, which attempt to adjust the inference from Cox's model in situations where ties occur. The predicted and conditional marginals are presented for both of these methods. The point estimates for the two types of marginals are derived using the pseudo-likelihood equations, and variance estimates are determined using Taylorized deviations. The variance estimates derived from the Taylorized deviations tend to be robust to cluster-correlated data. We will present analytic results comparing the variance estimates using the Taylorized deviations and the model-based (naïve) variance estimates. All calculations for the analytic examples are derived using new features in SUDAAN's SURVIVAL procedure.

Introduction

Calculation of predicted and conditional marginals is often desired when modeling data. These marginals present estimates of the response variable for different levels of a variable controlling for all other explanatory variables in the model. The results are presented on the scale of interest, thereby removing the difficulties of interpretation of regression coefficients and hazards ratios.

The marginals in Cox's proportional hazard model estimate the probability of an event not occurring (*i.e.* probability of surviving) given that the individual is in the population for the average length of time. If marginals are requested for all levels of a categorical variable, then individuals in the category with the highest percentage would be the least likely to undergo an event.

The predicted margins for Cox's model are derived in Korn and Graubard (1999). However, the formulation for the variances for these estimates is not fully presented.

The use of the Taylor series linearization allows calculation of variances for a variety of potential sampling frames. Other frequently used variance estimation methods, such as Balanced Repeated Replication (BRR) and Jackknife, often are restricted in their use to the types of surveys were they are appropriate. No such restrictions are applied to the Taylor series making them indispensable for variance calculation.

Cox Proportional Hazards – Taylor Series Designs

The primary focus of this paper is the calculation of the predicted and conditional marginals and their respective variances using the Taylorized deviations of these estimates. In order to derive these estimates, it is necessary to present the pseudo-likelihood for the proportional hazards model, then show the derivations of estimates and Taylorized deviations of the regression coefficients and the baseline hazards. We will show these estimates for Breslow's and Efron's methods of handling tied events.

The formulation of the likelihood for Cox's proportional hazards model can be found in Kalbfleisch and Prentice (2002). Since our results are specifically designed for correlated data and complex survey, we will only deal with the weighted likelihood given here:

$$L(\boldsymbol{\beta}) = \prod_{j=1}^{n} \left[\frac{\exp(\mathbf{x}'_j \hat{\boldsymbol{\beta}})}{\sum_{k=1}^{n} w_k I(t_{1k} < t_{2j} \le t_{2k}) \exp(\mathbf{x}'_k \hat{\boldsymbol{\beta}})} \right]^{\delta_j w_k} .$$

Here, t_{1k} indicates the time an individual enters the population and is at risk of failure. The value t_{2j} represents the time at which the individual undergoes an event or is censored. The parameter δ_j is an event indicator that is 1 if an event occurred at t_{2j} and 0 otherwise. The value w_j is the individual weight and $\mathbf{x}_j$ is a p-dimensional vector of covariates. The vector of repression coefficients is represented by $\boldsymbol{\beta}$.

First, we discuss the cases of no ties, or of Breslow's method of handling tied events. The following functions defined below will be used throughout the paper:

$$\hat{\mathbf{S}}_0(t,\hat{\boldsymbol\beta}) = \sum_{j=1}^{n} w_j I(t_{1j} < t \le t_{2j})\exp(\mathbf{x}'_j\hat{\boldsymbol\beta}) \ ,$$

$$\hat{\mathbf{S}}_1(t,\hat{\boldsymbol\beta}) = \sum_{j=1}^{n} w_j I(t_{1j} < t \le t_{2j})\exp(\mathbf{x}'_j\hat{\boldsymbol\beta})\mathbf{x}_j \ ,$$

$$\hat{\mathbf{S}}_2(t,\hat{\boldsymbol\beta}) = \sum_{j=1}^{n} w_j I(t_{1j} < t \le t_{2j})\exp(\mathbf{x}'_j\hat{\boldsymbol\beta})\mathbf{x}_j\mathbf{x}'_j \ .$$

Note that $\dfrac{\partial \mathbf{S}_0(t,\hat{\boldsymbol\beta})}{\partial\boldsymbol\beta} = \mathbf{S}_1(t,\hat{\boldsymbol\beta})$ and $\dfrac{\partial \mathbf{S}_1(t,\hat{\boldsymbol\beta})}{\partial\boldsymbol\beta} = \mathbf{S}_2(t,\hat{\boldsymbol\beta})$.
Further note that,
$\dfrac{\partial \mathbf{S}_0(t,\hat{\boldsymbol\beta})}{\partial w_i} = \mathbf{S}_1(t,\hat{\boldsymbol\beta})\dfrac{\partial\hat{\boldsymbol\beta}}{\partial w_i} + I(t_{1i} < t \le t_{2i})\exp(\mathbf{x}'_i\hat{\boldsymbol\beta})$ and that $\dfrac{\partial \mathbf{S}_1(t,\hat{\boldsymbol\beta})}{\partial w_i} = \mathbf{S}_2(t,\hat{\boldsymbol\beta})\dfrac{\partial\hat{\boldsymbol\beta}}{\partial w_i} + I(t_{1i} < t \le t_{2i})\exp(\mathbf{x}'_i\hat{\boldsymbol\beta})\mathbf{x}_i$.

Estimates of Coefficients in Cox's Proportional Hazards Model

Cox's proportional hazards model (Cox, 1972; Cox and Oakes, 1984) as detailed by Binder (1992) for sample surveys is:

$$L(\boldsymbol\beta) = \prod_{j=1}^{n}\left[\frac{h(t_{2j}\,|\,\mathbf{x}_j)}{h_0(t_{2j})\hat{\mathbf{S}}_0(t_{2j},\hat{\boldsymbol\beta})}\right]^{w_j\delta_j} = \prod_{j=1}^{n}\left[\frac{\exp(\mathbf{x}'_j\hat{\boldsymbol\beta})}{\hat{\mathbf{S}}_0(t_{2j},\hat{\boldsymbol\beta})}\right]^{w_j\delta_j}$$
.

The Cox model assumes the hazard function: $h(t|\mathbf{x}_j) = h_0(t)\exp(\mathbf{x}'_j\hat{\boldsymbol\beta})$, where $h_0(t)$ is the baseline hazard function at time t.

The log-likelihood is

$$LogL(\boldsymbol\beta) = \sum_{j=1}^{n} w_j\delta_j\left[\mathbf{x}'_j\hat{\boldsymbol\beta} - \log(\hat{\mathbf{S}}_0(t_{2j},\hat{\boldsymbol\beta}))\right] .$$

The parameters $\boldsymbol\beta$ are estimated by maximizing the weighted partial likelihood, and the variance-covariance matrix for $\hat{\boldsymbol\beta}$ is estimated by implicit Taylor series linearization. The resulting estimating equations (or score functions) are:

$$\mathbf{U}(\hat{\boldsymbol\beta}) = \frac{\partial LogL(\boldsymbol\beta)}{\partial\boldsymbol\beta} = \sum_{j=1}^{n} w_j\delta_j\left[\mathbf{x}_j - \frac{\hat{\mathbf{S}}_1(t_{2j},\hat{\boldsymbol\beta})}{\hat{\mathbf{S}}_0(t_{2j},\hat{\boldsymbol\beta})}\right] = 0$$

We can obtain the Taylor deviations of $\hat{\boldsymbol\beta}$ by differentiating the estimating equation with respect to w_i and then solving for $\dfrac{\partial\hat{\boldsymbol\beta}}{\partial w_i}$ which yields:

$$\frac{\partial\boldsymbol\beta}{\partial w_i} = -\mathbf{J}^{-1}\left[\begin{array}{l}\delta_i\left\{\mathbf{x}_i - \dfrac{\hat{\mathbf{S}}_1(t_{2i},\hat{\boldsymbol\beta})}{\hat{\mathbf{S}}_0(t_{2i},\hat{\boldsymbol\beta})}\right\} \\ -\displaystyle\sum_{j=1}^{n} w_j\delta_j\left[\frac{I(t_{1i} < t_{2j} \le t_{2i})\exp(\mathbf{x}'_i\hat{\boldsymbol\beta})}{\hat{\mathbf{S}}_0(t_{2j},\hat{\boldsymbol\beta})}\left\{\mathbf{x}_i - \frac{\hat{\mathbf{S}}_1(t_{2j},\hat{\boldsymbol\beta})}{\hat{\mathbf{S}}_0(t_{2j},\hat{\boldsymbol\beta})}\right\}\right]\end{array}\right]$$
.

The Taylor deviation of $\boldsymbol\beta$ with respect to i is $T_i(\boldsymbol\beta) = w_i\dfrac{\partial\boldsymbol\beta}{\partial w_i}$. The variance $\mathbf{V}(\hat{\boldsymbol\beta})$ is the variance of the sum of the Taylor deviations. This estimated robust variance-covariance matrix of $\hat{\boldsymbol\beta}$ has been derived by Binder (1992).

In order to illustrate Efron's method, the event times will be ordered. There are r unique event times in the t_{2j} set. If there are no ties, then $r=n$, otherwise $r < n$. Let $t_{2(h)}$ represent the ordered event times from the t_{2j} set such that $t_{2(1)} < t_{2(2)} < \ldots < t_{2(r)}$. Further let $n_{(h)}$ represent the total number of events plus censored observations at a given time $t_{2(h)}$ such that $\sum_{h=1}^{r} n_{(h)} = n$. Let the $n_{(h)}$ records that have an event or are censored at time $t_{2(h)}$ compose the end time set $E(t_{2(h)})$. Let $D(t_{2(h)})$ be a subset of $E(t_{2(h)})$ comprised of the $d_{(h)}$ observations that have a event at time $t_{2(h)}$. Let the vector of covariates for the records in $E(t_{2(h)})$ at a given ordered event time h be represented by $\mathbf{x}_{(h)1}$, $\mathbf{x}_{(h)2}$, … $\mathbf{x}_{(h)n(h)}$ and the weights represented by $w_{(h)1}$, $w_{(h)2}$, … $w_{(h)n(h)}$. Similar notation can be used for the events that compose $D(t_{2(h)})$.

In this section, we present the equations for the contribution to the likelihood from the ties, and their impact on the formulas already presented. Consider the case of tied records, each having an event at time $t_{2\,(h)}$.

The contribution to the likelihood from these events under Breslow is given by:

$$\mathbf{L}_{B_{(h)}} = \prod_{k=1}^{n_{(h)}} \left[\frac{\exp(\mathbf{x}'_{(h)_k} \hat{\boldsymbol{\beta}})}{\hat{\mathbf{S}}_{0B}(t_{2(h)}, \hat{\boldsymbol{\beta}})} \right]^{\delta_{(h)k} w_{(h)k}} = \prod_{m=1}^{d_{(h)}} \left[\frac{\exp(\mathbf{x}'_{(h)_m} \hat{\boldsymbol{\beta}})}{\hat{\mathbf{S}}_{0B}(t_{2(h)}, \hat{\boldsymbol{\beta}})} \right]^{w_{(h)m}}$$

where the exposure sum $\hat{\mathbf{S}}_{0B}(t_{2(h)}, \hat{\boldsymbol{\beta}})$ under Breslow's approximation includes all the individuals at risk at time $t_{2(h)}$. Note that all observations in $D(t_{2(h)})$ receive the same value for $\hat{\mathbf{S}}_{0B}(t_{2(h)}, \hat{\boldsymbol{\beta}})$ and that the δ_j are only needed when summing over the end time set. By using the same values for each event, Breslow's approximation overestimates the exposure. The exact method requires one to consider all possible permutations of $d_{(h)}$ events and take the average value; such an approach was deemed too computer intensive and not addressed in this paper. Efron has provided a simple approximation that produces results that closely approximate the exact method.

Efron's method is based on assuming that $(1/d_{(h)})$ of each of the $d_{(h)}$ events (that is events with weight equal to $w_{(h)_k} / d_{(h)}$) occurred at time point [$t_{2(h)k} = t_{2(h)} - (k-1)\varepsilon$], for k = 1, 2, ..., $d_{(j)}$; where ε is a very small quantity of time such that none of these time points occur before the previous events. From this, a complete ordering of all death events is $t_{2(1)1} < t_{2(1)2} \ldots < t_{2(1)d(1)} < t_{2(2)1} \ldots < t_{2(2)d(2)} \ldots < t_{2(r)1} \ldots < t_{2(r)d(r)}$. By considering only the observations in $D(t_{2(h)})$ at the h^{th} ordered event time, Efron's partial likelihood for this set is

$$\begin{aligned}
&\mathbf{L}_{E_{(h)}}(D(t_{2(h)})) \\
&= \prod_{m=1}^{d_{(h)}} \frac{\left[\exp(\mathbf{x}'_{(h)m} \hat{\boldsymbol{\beta}})\right]^{w_{(h)m}}}{\left[\mathbf{S}_{0B}(t_{2(h)}, \hat{\boldsymbol{\beta}}) - \dfrac{m-1}{d_{(h)}} \displaystyle\sum_{k=1}^{d_{(h)}} w_{(h)_k} \exp(\mathbf{x}'_{(h)_k} \hat{\boldsymbol{\beta}}) \right]^{\bar{w}_{(h)}}} \\
&= \prod_{m=1}^{d_{(h)}} \frac{\left[\exp(\mathbf{x}'_{(h)m} \hat{\boldsymbol{\beta}})\right]^{w_{(h)m}}}{\left[\mathbf{S}_{0E}(t_{2(h)}, \hat{\boldsymbol{\beta}}, m)\right]^{\bar{w}_{(h)}}}
\end{aligned}$$

where $\bar{w}_{(h)}$ is the average of the weights of all observations in $D(t_{2(h)})$. The pseudo-likelihood encompassing all ordered event times is

$$L_E(D(t_{2(h)})) = \prod_{h=1}^{r} L_{E_{(h)}}(D(t_{2(h)})) \text{ .}$$

A few definitions are needed to extend the likelihood to the full data set. The subscript (h) refers to observations at the h^{th} ordered event time. The subscript j refers to the j^{th} observation in the data set. The term $d_{(h)}$ is the number of observations that have an event a the h^{th} ordered event time. The term d_j is the number of observations that share the same event time as observation j. A similar definition applies to the mean weights $\bar{w}_j$ and $\bar{w}_{(h)}$. Any given observation falls into a single event set. In order to avoid further complicating the notation, the event set will be implied. Any observation j with event time t_{2j} falls into exactly one of the $E(t_{2(h)})$ end time sets. Thus, the expression $\mathbf{S}_{0E}(t_{2j}, \hat{\boldsymbol{\beta}}, m)$ is equivalent to $\mathbf{S}_{0E}(t_{2(h)}, \hat{\boldsymbol{\beta}}, m)$ provided $j \in E(t_{2(h)})$ which implies $t_{2j} = t_{2(h)}$. The assumption that $j \in D(t_{2(h)}) \in E(t_{2(h)})$ is implicit and therefore is not stated in these expressions. In the following expressions, the sums can be taken over the entire set of observations or can be taken over the event set that contains the j^{th} observation. Therefore, we define the following sums in the presence of ties analogous to those defined above for the case with no ties:

$$\mathbf{S}_{0E}(t_{2j}, \hat{\boldsymbol{\beta}}, m) = \hat{\mathbf{S}}_0(t_{2j}, \hat{\boldsymbol{\beta}}) - \frac{m-1}{d_j} \sum_{k=1}^{n} w_k I(t_{2k} = t_{2j}) \delta_k \exp(\mathbf{x}'_k \hat{\boldsymbol{\beta}})$$

is equivalent to

$$\mathbf{S}_{0E}(t_{2(h)}, \hat{\boldsymbol{\beta}}, m) = \hat{\mathbf{S}}_0(t_{2(h)}, \hat{\boldsymbol{\beta}}) - \frac{m-1}{d_{(h)}} \sum_{k=1}^{d_{(h)}} w_k \exp(\mathbf{x}'_k \hat{\boldsymbol{\beta}}) \text{ .}$$

While not presented here, the following equations can also be expressed in a similar manner:

$$\mathbf{S}_{1E}(t_{2j}, \hat{\boldsymbol{\beta}}, m) = \hat{\mathbf{S}}_1(t_{2j}, \hat{\boldsymbol{\beta}}) - \frac{m-1}{d_{(h)}} \sum_{k=1}^{d_{(h)}} w_k \exp(\mathbf{x}'_k \hat{\boldsymbol{\beta}}) \mathbf{x}_k$$

$$\mathbf{S}_{2E}(t_{2j}, \hat{\boldsymbol{\beta}}, m) = \hat{\mathbf{S}}_2(t_{2j}, \hat{\boldsymbol{\beta}}) - \frac{m-1}{d_j} \sum_{k=1}^{n} w_k I(t_{2k} = t_{2j}) \delta_k \exp(\mathbf{x}'_k \hat{\boldsymbol{\beta}}) \mathbf{x}_k \mathbf{x}'_k$$

Note that

$$\frac{\partial \mathbf{S}_{0E}(t_{2j}, \hat{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}} = \mathbf{S}_{1E}(t_{2j}, \hat{\boldsymbol{\beta}})$$

and

$$\frac{\partial \mathbf{S}_{1E}(t_{2j}, \hat{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}} = \mathbf{S}_{2E}(t_{2j}, \hat{\boldsymbol{\beta}})$$

and that

$$\frac{\partial \mathbf{S}_{0E}(t_{2j},\hat{\boldsymbol\beta},m)}{\partial w_i} = \mathbf{S}_{1E}(t_{2j},\hat{\boldsymbol\beta},m)\frac{\partial\hat{\boldsymbol\beta}}{\partial w_i} + \left[I(t_{1i}<t_{2j}\le t_{2i}) - \frac{m-1}{d}I(t_{2i}=t_{2j})\delta_i\right]\exp(\mathbf{x}_i'\hat{\boldsymbol\beta})$$

and

$$\frac{\partial \mathbf{S}_{1E}(t_{2j},\hat{\boldsymbol\beta},m)}{\partial w_i} = \mathbf{S}_{2E}(t_{2j},\hat{\boldsymbol\beta},m)\frac{\partial\hat{\boldsymbol\beta}}{\partial w_i} + \left[I(t_{1i}<t_{2j}\le t_{2i}) - \frac{m-1}{d}I(t_{2i}=t_{2j})\delta_i\right]\exp(\mathbf{x}_i'\hat{\boldsymbol\beta})\mathbf{x}_i$$

The log-likelihood in this case is

$$LogL_E(\boldsymbol\beta) = \sum_{j=1}^{n}\delta_j w_j \mathbf{x}_j'\hat{\boldsymbol\beta} - \sum_{h=1}^{r}\overline{w}_{(h)}\sum_{m=1}^{d_{(h)}}\log(\hat{\mathbf{S}}_{0E}(t_{2(h)},\hat{\boldsymbol\beta},m))$$

The estimating equations (score functions) are:

$$\mathbf{U}_E(\hat{\boldsymbol\beta}) = \frac{\partial LogL_E(\boldsymbol\beta)}{\partial\boldsymbol\beta} = \sum_{j=1}^{n}\delta_j w_j \mathbf{x}_j - \sum_{h=1}^{r}\overline{w}_{(h)}\sum_{m=1}^{d_{(h)}}\frac{\hat{\mathbf{S}}_{1E}(t_{2(h)},\hat{\boldsymbol\beta},m)}{\hat{\mathbf{S}}_{0E}(t_{2(h)},\hat{\boldsymbol\beta},m)} = 0$$

Once again we can obtain the Taylor deviation of $\hat{\boldsymbol\beta}$ by differentiating the estimating equation with respect to w_i and then solving for $\frac{\partial\hat{\boldsymbol\beta}}{\partial w_i}$:

$$\frac{\partial\boldsymbol\beta}{\partial w_i} = -\mathbf{J}_E^{-1}\left[\delta_i\left\{\mathbf{x}_i - \sum_{m=1}^{d_{(h)}}\frac{\hat{\mathbf{S}}_{1E}(t_{2i},\hat{\boldsymbol\beta},m)}{\hat{\mathbf{S}}_{0E}(t_{2i},\hat{\boldsymbol\beta},m)}\right\} - \sum_{h=1}^{r}\overline{w}_{(h)}\sum_{m=1}^{d_{(h)}}\left[\frac{\left\{I(t_{1i}<t_{2(h)}\le t_{2i}) - \frac{m-1}{d}I(t_{2i}=t_{2(h)})\right\}\exp(\mathbf{x}_i'\hat{\boldsymbol\beta})}{\hat{\mathbf{S}}_{0E}(t_{2i},\hat{\boldsymbol\beta},m)}\left\{\mathbf{x}_i - \frac{\hat{\mathbf{S}}_{1E}(t_{2(h)},\hat{\boldsymbol\beta},m)}{\hat{\mathbf{S}}_{0E}(t_{2(h)},\hat{\boldsymbol\beta},m)}\right\}\right]\right]$$

where

$$\mathbf{J}_E = \frac{\partial\mathbf{U}_E(\hat{\boldsymbol\beta})}{\partial\boldsymbol\beta} = \sum_{h=1}^{r}\overline{w}_{(h)}\sum_{m=1}^{d_{(h)}}\left[\frac{\hat{\mathbf{S}}_{1E}(t_{2(h)},\hat{\boldsymbol\beta},m)\hat{\mathbf{S}}_{1E}{}'(t_{2(h)},\hat{\boldsymbol\beta},m)}{\hat{\mathbf{S}}_{0E}^{2}(t_{2(h)},\hat{\boldsymbol\beta},m)} - \frac{\hat{\mathbf{S}}_{2E}(t_{2(h)},\hat{\boldsymbol\beta},m)}{\hat{\mathbf{S}}_{0E}(t_{2(h)},\hat{\boldsymbol\beta},m)}\right].$$

The Baseline Hazard Function

Breslow (1972) has presented a simple estimator for the baseline hazard function by equating the actual number of events in the interval (0, *t*] to the expected number of events up to time *t* for all subjects. The resulting estimator is

$$\hat{H}_0(t) = \sum_{j=1}^{n}\frac{\delta_j w_j I(t_{2j}\le t)}{\sum_{k=1}^{n} w_k I(t_{1k}<t_{2j}\le t_{2k})\exp(\hat{\boldsymbol\beta}'\mathbf{x})}$$

We evaluate this estimate at $\boldsymbol\beta = \hat{\boldsymbol\beta}$ obtained by maximizing the partial likelihood function. In order to describe the behavior of the hazard function in case of ties, it is helpful to see that the estimated hazard function is a step function with a jump at each time point for an event, and the size of the jump $\Delta\hat{H}_0(t_{2j})$ with j^{th} observation with an event at time t_{2j} is

$$\Delta\hat{H}_0(t_{2j}) = \frac{\delta_j w_j}{\hat{\mathbf{S}}_0(t_{2j},\hat{\boldsymbol\beta})}.$$

In case of ties with $n_{(h)}$ end time events at time *h*, the total jump with Breslow's approximation is

$\Delta\hat{H}_{0B}(t_{2j}=t_{2(h)})=\dfrac{\sum_{k=1}^{n_{(h)}}\delta_{(h)_k}w_{(h)_k}}{\hat{\mathbf{S}}_0(t_{2(h)},\hat{\boldsymbol{\beta}})}$, where the sum in the numerator is over $E(t_{2(h)})$ at time $t_{2(h)}$. If Efron's approximation is used, then the expression for the total jump is:

$$\Delta\hat{H}_{0E}(t_{2j}=t_{2(h)})=\delta_j\overline{w}_{(h)}\sum_{m=1}^{d_{(h)}}\frac{1}{\hat{\mathbf{S}}_{0E}(t_{2(h)},\hat{\boldsymbol{\beta}},m)},$$

where the sum is over the event set $D(t_{2(h)})$.

The Taylorized deviation of $\hat{H}_0(t)$ (no ties) can be computed as:

$$T_i\left[\hat{H}_0(t)\right]=T_i\left[\sum_{j=1}^{n}I(t_{2j}\le t)\Delta\hat{H}_0(t_{2j})\right]=\sum_{j=1}^{n}I(t_{2j}\le t)T_i\left[\Delta\hat{H}_0(t_{2j})\right].$$

The Taylorized deviation of $\Delta\hat{H}_0(t_{2j})$ with respect to the i^{th} observation with a single event is:

$$T_i[\Delta\hat{\mathrm{H}}_0(\mathrm{t}_{2\mathrm{j}})]=\frac{\delta_i w_i I(i=j)}{\hat{\mathbf{S}}_0(t_{2i},\hat{\boldsymbol{\beta}})}$$
$$-\Delta\hat{\mathrm{H}}_0(\mathrm{t}_{2\mathrm{j}})\left[\frac{w_i I(t_{1i}<t_{2j}\le t_{2i})\exp(\mathbf{x}_i'\hat{\boldsymbol{\beta}})+\hat{\mathbf{S}}_1'(t_{2j},\hat{\boldsymbol{\beta}})T_i[\hat{\boldsymbol{\beta}}]}{\hat{\mathbf{S}}_0(t_{2j},\hat{\boldsymbol{\beta}})}\right]$$

The corresponding expression in case of ties with Breslow's approximation is:

$$T_i[\Delta\hat{H}_{0B}(t_{2j}=t_{2(h)})]=$$

$$w_i\frac{\hat{\mathbf{S}}_0(t_{2(h)},\hat{\boldsymbol{\beta}})\left[\sum_{k=1}^{n_{(h)}}\delta_{(h)_k}I(i=(h)_k)\right]-\left[\sum_{k=1}^{n_{(h)}}\delta_{(h)_k}w_{(h)_k}\right]\left[I(t_{1i}<t_{2(h)}\le t_{2i})\exp(\mathbf{x}_i'\hat{\boldsymbol{\beta}})+\hat{\mathbf{S}}_1'(t_{2(h)},\hat{\boldsymbol{\beta}})\frac{\partial\hat{\boldsymbol{\beta}}}{\partial w_i}\right]}{\hat{\mathbf{S}}_0^2(t_{2(j)},\hat{\boldsymbol{\beta}})}$$
$$=\frac{\delta_i w_i I(t_{2i}=t_{2(h)})}{\hat{\mathbf{S}}_0(t_{2(h)},\hat{\boldsymbol{\beta}})}-\Delta\hat{H}_{0B}(t_{2j}=t_{2(h)})\frac{\left[w_i I(t_{1i}<t_{2(h)}\le t_{2i})\exp(\mathbf{x}_i'\hat{\boldsymbol{\beta}})+\hat{\mathbf{S}}_1'(t_{2(h)},\hat{\boldsymbol{\beta}})T_i[\hat{\boldsymbol{\beta}}]\right]}{\hat{\mathbf{S}}_0(t_{2(h)},\hat{\boldsymbol{\beta}})}$$

For Efron's approximation in case of ties the expression is:

$$T_i[\Delta\hat{H}_{0E}(t_{2j}=t_{2(h)})]$$
$$=\left[\frac{\delta_j w_i I(i\in D(t_{2(h)}))}{d_{(h)}}\sum_{k=1}^{d_{(h)}}\frac{1}{\hat{\mathbf{S}}_{0E}(t_{2(h)},\hat{\boldsymbol{\beta}},k)}-\delta_j\overline{w}_{(h)}\sum_{k=1}^{d_{(h)}}\frac{\hat{\mathbf{S}}_{1E}(t_{2(h)},\hat{\boldsymbol{\beta}},k)T_i[\hat{\boldsymbol{\beta}}]+w_i\exp(\mathbf{x}_i'\hat{\boldsymbol{\beta}})\left\{I(t_{1i}<t_{2(h)}\le t_{2i})-\left\{\frac{(k-1)I(t_{2i}=t_{2(h)})\delta_i}{d_{(h)}}\right\}\right\}}{\hat{\mathbf{S}}_{0E}^2(t_{2(h)},\hat{\boldsymbol{\beta}},k)}\right]$$

For computing the Taylorized deviations for the predicted and conditional marginals we will need the Taylorized deviation of $\hat{H}_0(t_{2j})-\hat{H}_0(t_{1j})$. Note that for a single observation, j, this difference is

$$\hat{H}_0(t_{2j})-\hat{H}_0(t_{1j})=\sum_{k=1}^{n}I(t_{1j}<t_{2k}\le t_{2j})\Delta\hat{H}_0(t_{2k})$$

Thus

$$T_i\left[\hat{H}_0(t_{2p})-\hat{H}_0(t_{1p})\right]=\sum_{j=1}^{n}I(t_{1p}<t_{2j}\le t_{2p})T_i\left[\Delta\hat{H}_0(t_{2j})\right]$$

.

Predicted Marginals

Let A be a categorical variable and let a represent a given level of A. An estimate of the predicted marginal survival for a given level of A is equivalent to direct standardization. Specifically, it is the expected survival for a hypothetical population of individuals such that every individual has *A*=*a* with all other characteristics the same as those in the observed population. The predicted marginal survival for *A*=*a* is defined as

$$\hat{\mathbf{P}}_A = \frac{1}{w_+}\sum_{p=1}^{n} w_p \exp\left(-\left\{\hat{H}_0(t_{2p}) - \hat{H}_0(t_{1p})\right\}\exp(x^*_p\hat{\boldsymbol{\beta}})\right)$$
$$= \frac{1}{w_+}\sum_{p=1}^{n} w_p K_p$$

where

$$w_+ = \sum_{p=1}^{n} w_p \ ,$$
$$K_p = \exp\left(-\left\{\hat{H}_0(t_{2p}) - \hat{H}_0(t_{1p})\right\}\exp(x^*_p\hat{\boldsymbol{\beta}})\right),$$

and the vector $\mathbf{x}^*_j$ is the vector $\mathbf{x}_j$ except that the characteristics specified for the predicted marginal are assigned the specified value; *e.g.* the vector $\mathbf{x}^*_j$ for every observation is assigned the value *A*=*a*.

For estimating the variance of the predicted marginal we need to derive the Taylorized deviation of $\hat{\mathbf{P}}_A$:

$$T_i\left[\hat{P}_A\right] = \frac{w_i}{w_+}\left[K_i - \hat{P}_A\right] + \frac{1}{w_+}\sum_{p=1}^{n} w_p T_i[K_p]$$

and

$$T_i[K_p] = -K_p\begin{bmatrix} T_i\left[\hat{H}_0(t_{2p}) - \hat{H}_0(t_{1p})\right]\exp(\mathbf{x}^*_p\hat{\boldsymbol{\beta}}) \\ + \left\{\hat{H}_0(t_{2p}) - \hat{H}_0(t_{1p})\right\}\exp(\mathbf{x}^*_p\hat{\boldsymbol{\beta}})\mathbf{x}^*_p T_i[\boldsymbol{\beta}] \end{bmatrix}$$

Note that in the last equation the terms $\mathbf{x}^*_p\hat{\boldsymbol{\beta}}$ and $\mathbf{x}^*_p \frac{\partial\hat{\boldsymbol{\beta}}}{\partial w_i}$ are scalar products and that $\hat{\boldsymbol{\beta}}$ and $\frac{\partial\hat{\boldsymbol{\beta}}}{\partial w_i}$ are evaluated at observed $\mathbf{x}_p$ and not at hypothetical $\mathbf{x}^*_p$.

If we now write out the complete Taylorized deviation for $\hat{\mathbf{P}}_A$ we see that there are many nested sums over the entire data set which must be computed. We show next the full details along with a rearrangement which will be most efficient for calculations.

Step 1:

$$T_i\left[\hat{P}_A\right] = \frac{w_i}{w_+}\left[K_i - \hat{P}_A\right] + \frac{1}{w_+}\sum_{p=1}^{n} w_p T_i[K_p]$$

Step 2:

$$T_i[\hat{P}_A] = \frac{w_i}{w_+}[K_i - \hat{P}_A]$$
$$-\frac{1}{w_+}\sum_{p=1}^{n} w_p K_p\begin{bmatrix} T_i\left[\hat{H}_0(t_{2p}) - \hat{H}_0(t_{1p})\right]\exp(\mathbf{x}^*_p\hat{\boldsymbol{\beta}}) \\ + \left\{\hat{H}_0(t_{2p}) - \hat{H}_0(t_{1p})\right\}\exp(\mathbf{x}^*_p\hat{\boldsymbol{\beta}})\mathbf{x}^*_p T_i[\hat{\boldsymbol{\beta}}] \end{bmatrix}$$

Step 3:

$$T_i\left[\hat{P}\right] = \frac{w_i}{w_+}\left[K_i - \hat{P}\right]$$
$$-\frac{1}{w_+}\left\{\sum_{p=1}^{n} w_p K_p \exp(\mathbf{x}^*_p\hat{\boldsymbol{\beta}})\left\{\hat{H}_0(t_{2p}) - \hat{H}_0(t_{1p})\right\}\mathbf{x}^*_p\right\} T_i[\boldsymbol{\beta}]$$
$$-\frac{1}{w_+}\sum_{p=1}^{n} w_p K_p \exp(\mathbf{x}^*_p\hat{\boldsymbol{\beta}}) T_i\left[\hat{H}_0(t_{2p}) - \hat{H}_0(t_{1p})\right]$$

The sums in the third term can be rearranged as follows:

Step 4:

$$\frac{1}{w_+}\sum_{p=1}^{n} w_p K_p \exp(\mathbf{x}^*_p\hat{\boldsymbol{\beta}}) T_i\left[\hat{H}_0(t_{2p}) - \hat{H}_0(t_{1p})\right]$$
$$= \frac{1}{w_+}\sum_{p=1}^{n} w_p K_p \exp(\mathbf{x}^*_p\hat{\boldsymbol{\beta}})\sum_{j=1}^{n} I(t_{1p} < t_{2j} \le t_{2p}) T_i\left[\Delta\hat{H}_0(t_{2j})\right]$$
$$= \frac{1}{w_+}\sum_{j=1}^{n} T_i\left[\Delta\hat{H}_0(t_{2j})\right]\sum_{p=1}^{n} w_p K_p \exp(\mathbf{x}^*_p\hat{\boldsymbol{\beta}}) I(t_{1p} < t_{2j} \le t_{2p})$$
$$= \frac{1}{w_+}\sum_{j=1}^{n} T_i\left[\Delta\hat{H}_0(t_{2j})\right] M_j$$

where

$$M_j = \sum_{p=1}^{n} w_p K_p \exp(\mathbf{x}^*_p\hat{\boldsymbol{\beta}}) I(t_{1p} < t_{2j} \le t_{2p})$$

The variance of the predicted marginal survival can be estimated using the above Taylorized deviations. The estimate for the contrasted predicted marginal survival is a linear function of several predicted marginals:

$$L(P_1, P_2, ..., P_K) = \sum_{k=1}^{K} l_k P_k$$

and the Taylor deviation of the contrast can be estimated by taking the same linear function of the Taylor deviation of the predicted marginals:

$$T\,[L] = T_i\left[\sum_{k=1}^{K} l_k P_k\right] = \sum_{k=1}^{k} l_k T_i[P_k]\,.$$

Conditional Marginals

An estimate of the conditional marginal survival for category A=a is equivalent to the expected survival for a person with A=a and with all other characteristics equal to the average for those variables in the observed population. The conditional marginal survival for A=a is defined as

$$\hat{C}_A = \exp\left(-\frac{1}{w_+}\sum_{p=1}^{n} w_p\{\hat{H}_0(t_{2p}) - \hat{H}_0(t_{1p})\}\exp(\overline{\mathbf{x}}^*\hat{\boldsymbol{\beta}})\right),$$

where $\overline{\mathbf{x}}^*$ is the weighted average of $\mathbf{x}_p^*$ defined for predicted marginals

$$\overline{\mathbf{x}}^* = \frac{1}{w_+}\sum_{p=1}^{n} w_p \mathbf{x}_p^*\,.$$

The conditional survival can then be written as

$$\hat{C}_A = \exp\left(-\overline{H}_0\exp(\overline{\mathbf{x}}^*\hat{\boldsymbol{\beta}})\right),$$

where $\overline{H}_0 = \frac{1}{w_+}\sum_{p=1}^{n} w_p\{\hat{H}_0(t_{2p}) - \hat{H}_0(t_{1p})\}$ is the average baseline hazard experienced by all individuals.

The Taylorized deviation for the conditional marginal survival is

$$T_i\left[\hat{C}_A\right] = -\hat{C}_A\left[\overline{H}_0\overline{\mathbf{x}}^*\exp(\overline{\mathbf{x}}^*\hat{\boldsymbol{\beta}})T_i[\hat{\boldsymbol{\beta}}] + \exp(\overline{\mathbf{x}}^*\hat{\boldsymbol{\beta}})T_i\left[\overline{H}_0\right]\right]$$

$$= -\hat{C}_A\left[\begin{array}{l}\overline{H}_0\overline{\mathbf{x}}^*\exp(\overline{\mathbf{x}}^*\hat{\boldsymbol{\beta}})T_i[\hat{\boldsymbol{\beta}}] \\ +\exp(\overline{\mathbf{x}}^*\hat{\boldsymbol{\beta}})\left\{\frac{-w_i}{w_+^2}\sum_{p=1}^{n} w_p\{\hat{H}_0(t_{2p}) - \hat{H}_0(t_{1p})\} + \frac{w_i}{w_+}\{\hat{H}_0(t_{2p}) - \hat{H}_0(t_{1p})\} + \frac{1}{w_+}\sum_{p=1}^{n} w_p\sum_{j=1}^{n} I(t_{1p} < t_{2j} \le t_{2p})T_i\left[\Delta\hat{H}_0(t_{2j})\right]\right\}\end{array}\right]$$

$$= \frac{-\hat{C}_A\exp(\overline{\mathbf{x}}^*\hat{\boldsymbol{\beta}})}{w_+}\left[w_+\overline{H}_0\overline{\mathbf{x}}^*T_i[\hat{\boldsymbol{\beta}}] - w_i\overline{H}_0 + w_i\left\{\hat{H}_0(t_{2i}) - \hat{H}_0(t_{1i})\right\} + \sum_{p=1}^{n} w_p\sum_{j=1}^{n} I(t_{1p} < t_{2j} \le t_{2p})T_i\left[\Delta\hat{H}_0(t_{2j})\right]\right]$$

Note that only the $\overline{\mathbf{x}}^*$ is treated as constant in the calculation of the Taylor deviations.

We have already specified the remaining Taylor deviations above. The variance of the conditional marginal is estimated from these Taylor deviations. The derivation of the Taylor deviations for the contrasts of the conditional marginals is similar to that for contrasts of predicted marginals.

References

Breslow, N.E. (1972), Discussion of "Covariance analysis of censored survival data," by D. R. Cox. *Journal of the Royal Statistical Society*, Series B, 34, 216-217.

Breslow, N.E. (1974). “Covariance analysis of censored survival data.” *Biometrics*, 30, 89-90.

Binder, D.A. (1992). “Fitting Cox’s proportional hazards models from survey data.” *Biometrika*, 79, 1, 139-147.

Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society*, Series B, 34, 187-220.

Cox, D. R. and D. Oakes (1984). Analysis of Survival Data.. London: Chapman and Hall.

Kalbfleisch, J.D. and Prentice, R.L. (2002). The Statistical Analysis of Failure Time Data. Second Edition. New York: Wiley.

Korn, E.L., and B.I. Graubard (1999). “Predictive margins for survey data.” *Biometrics* 55, 652-659.

Korn, E. L. and B. I. Graubard (1999). Analysis of Health Surveys. New York: Wiley.