Validation for the Cox Model

Terry M. Therneau
Mayo Foundation

Oct 2010

It is useful to have a set of test data where the results have been worked out in detail, both to illuminate the computations and to form a test case for software programs. The data sets below are quite simple, but have proven useful in this regard.

1 Test data 1

In this data set x is a 0/1 treatment covariate, with $n = 6$ subjects. There is one tied death time, one time with a death and a censoring, one with only a death, and one with only a censoring. (This is as small as a data set can be and still cover the four cases.) Let $r = \exp(\beta)$ be the risk score for a subject with $x = 1$. Table **??** shows the data set along with the mean and increment to the hazard at each point.

1.1 Breslow estimates

The loglikelihood has a term for each event; each term is the log of the ratio of the score for the subject who had an event over the sum of scores for

			$\bar{x}(t)$		$d\hat{\Lambda}_0(t)$	
Time	Status	x	Breslow	Efron	Breslow	Efron
1	1	1	$r/(r+1)$	$r/(r+1)$	$1/(3r+3)$	$1/(3r+3)$
1	0	1				
6	1	1	$r/(r+3)$	$r/(r+3)$	$2/(r+3)$	$1/(r+3)$
6	1	0		$r/(r+5)$		$2/(r+5)$
8	0	0				
9	1	0	0	0	1	1

Table 1: Test data 1

those who did not.

$$\begin{aligned} LL &= \{\beta - \log(3r+3)\} + \{\beta - \log(r+3)\} + \{0 - \log(r+3)\} + \{0-0\} \\ &= 2\beta - \log(3r+3) - 2\log(r+3). \end{aligned}$$

$$\begin{aligned} U &= \left(1 - \frac{r}{r+1}\right) + \left(1 - \frac{r}{r+3}\right) + \left(0 - \frac{r}{r+3}\right) + (0-0) \\ &= \frac{-r^2+3r+6}{(r+1)(r+3)}. \end{aligned}$$

$$\begin{aligned} \mathcal{I} &= \left\{\frac{r}{r+1} - \left(\frac{r}{r+1}\right)^2\right\} + 2\left\{\frac{r}{r+3} - \left(\frac{r}{r+3}\right)^2\right\} + (0-0) \\ &= \frac{r}{(r+1)^2} + \frac{6r}{(r+3)^2}. \end{aligned}$$

The actual solution corresponds to $U(\beta) = 0$, which from the quadratic formula is $r = (1/2)(3+\sqrt{33}) \approx 4.372281$, or $\hat{\beta} = \log(r) \approx 1.475285$. Then

$$\begin{array}{ll} LL(0) = -4.564348 & LL(\hat{\beta}) = -3.824750 \\ U(0) = 1 & U(\hat{\beta}) = 0 \\ \mathcal{I}(0) = 5/8 = 0.625 & \mathcal{I}(\hat{\beta}) = 0.634168\,. \end{array}$$

Newton–Raphson iteration has increments of $-\mathcal{I}^{-1}U$. Starting with the usual initial estimate of $\beta = 0$, the N–R iterates are 0, 8/5, 1.4727235, 1.4752838, 1.4752849, ... S considers the algorithm to have converged after three iterations, SAS after four (using the default convergence criteria in each package).

The martingale residuals are a simple function of the cumulative hazard, $M_i = \delta_i - r\hat{\Lambda}(t_i)$.

Subject	Λ_0	$\widehat{M}(0)$	$\widehat{M}(\hat{\beta})$
1	$1/(3r+3)$	5/6	.728714
2	$1/(3r+3)$	–1/6	–.271286
3	$1/(3r+3) + 2/(r+3)$	1/3	–.457427
4	$1/(3r+3) + 2/(r+3)$	1/3	.666667
5	$1/(3r+3) + 2/(r+3)$	–2/3	–.333333
6	$1/(3r+3) + 2/(r+3) + 1$	–2/3	–.333333

The score residual L_i can be calculated from the entries in Table **??**. For subject number 3, for instance, we have

$$L_3 = \int_0^6 \{1 - \bar{x}(t)\} d\widehat{M}_3(t)$$

$$= \left(1 - \frac{r}{r+1}\right)\frac{r}{3r+3} + \left(1 - \frac{r}{r+3}\right)\left(1 - \frac{2r}{r+3}\right).$$

Let $a = (r+1)(3r+3)$ and $b = (r+3)^2$; then the residuals are as follows.

Subject	L	$L(0)$	$L(\hat{\beta})$
1	$(2r+3)/a$	5/12	.135643
2	$-r/a$	–1/12	–.050497
3	$-r/a + 3(3-r)/b$	7/24	–.126244
4	$r/a - r(r+1)/b$	–1/24	–.381681
5	$r/a + 2r/b$	5/24	.211389
6	$r/a + 2r/b$	5/24	.211389

The Schoenfeld residuals are defined at the three unique death times, and have values of $1 - r/(r+1) = 1/(r+1)$, $\{1 - r/(r+3)\} + \{0 - r/(r+3)\} = (3-r)/(3+r)$, and 0 at times 1, 6, and 9, respectively. For convenience in plotting and use, however, the programs return one residual for each event rather than one per unique event time. The two values returned for time 6 are $3/(r+3)$ and $-r/(r+3)$.

The Nelson–Aalen estimate of the hazard is closely related to the Breslow approximation for ties. The baseline hazard is shown as the column Λ_0 above. The hazard estimate for a subject with covariate x_i is $\Lambda_i(t) = \exp(x_i\beta)\Lambda_0(t)$ and the survival estimate is $S_i(t) = \exp(-\Lambda_i(t))$.

The variance of the cumulative hazard is the sum of two terms. Term 1 is a natural extension of the Nelson–Aalen estimator to the case where there are weights. It is a running sum, with an increment at each death time of $dN(t)/(\sum Y_i(t)r_i(t))^2$. For a subject with covariate x_i this term is multiplied by $[\exp(x_i\beta)]^2$.

The second term is $d\mathcal{I}^{-1}d'$, where $\mathcal{I}$ is the variance–covariance matrix of the Cox model, and d is a vector. The second term accounts for the fact that the weights themselves have a variance; d is the derivative of $S(t)$ with respect to β and can be formally written as

$$\exp(x\beta)\int_0^t (\bar{x}(s) - x_i)d\hat{\Lambda}_0(s)\,.$$

This can be recognized as -1 times the score residual process for a subject with x_i as covariates and no events; it measures leverage of a particular observation on the estimate of β. It is intuitive that a small score residual — an obs with such covariates has little influence on β — results in a small added variance; that is, β has little influence on the estimated survival.

Time	Term 1
1	$1/(3r+3)^2$
6	$1/(3r+3)^2 + 2/(r+3)^2$
9	$1/(3r+3)^2 + 2/(r+3)^2 + 1/1^2$

Time	d
1	$(r/(r+1)) * 1/(3r+3)$
6	$(r/(r+1)) * 1/(3r+3) + (r/(r+3)) * 2/(r+3)$
9	$(r/(r+1)) * 1/(3r+3) + (r/(r+3)) * 2/(r+3) + 0 * 1$

For $\beta = 0$, $x = 0$:

Time	Variance		
1	$1/36$	$+ 1.6 * (1/12)^2$	$= 7/180$
6	$(1/36 + 2/16)$	$+ 1.6 * (1/12 + 2/16)^2$	$= 2/9$
9	$(1/36 + 2/16 + 1)$	$+ 1.6 * (1/12 + 2/16 + 0)^2$	$= 11/9$

For $\beta = 1.4752849$, $x = 0$

Time	Variance		
1	0.0038498	+ .004021	= 0.007871
2	0.040648	+ .0704631	= 0.111111
4	1.040648	+ .0704631	= 1.111111

1.2 Efron approximation

The Efron approximation [**?**] differs from the Breslow only at day 6, where two deaths occur. A useful way to think about the approximation is this: assume that if the data had been measured with higher accuracy that the deaths would not have been tied, that is two cases died on day 6 but they did not perish at the same instant on that day. There are thus two separate events on day 6. Four subjects were alive and at risk for the first of the events. Three subjects were at risk for the second event, either subjects 3, 5, and 6 or subjects 2, 5, and 6, but we do not know which. In some sense then, subjects 3 and 4 each have ".5" probability of being at risk for the second event at time $2 + \epsilon$. In the computation, we treat the two deaths as two separate times (two terms in the loglik), with subjects 3 and 4 each having a case weight of 1/2 for the second event. The mean covariate for the second event is then

$$\frac{1 * r/2 + 0 * 1/2 + 0 * 1 + 0 * 1}{r/2 + 1/2 + 1 + 1} = \frac{r}{r+5}$$

and the main quantities are

$$\begin{aligned} LL &= \{\beta - \log(3r+3)\} + \{\beta - \log(r+3)\} + \{0 - \log(r/2+5/2)\} + \{0-0\} \\ &= 2\beta - \log(3r+3) - \log(r+3) - \log(r/2+5/2) \end{aligned}$$

$$\begin{aligned} U &= \left(1 - \frac{r}{r+1}\right) + \left(1 - \frac{r}{r+3}\right) + \left(0 - \frac{r}{r+5}\right) + (0-0) \\ &= \frac{-r^3 + 23r + 30}{(r+1)(r+3)(r+5)} \end{aligned}$$

$$\begin{aligned} I &= \left\{\frac{r}{r+1} - \left(\frac{r}{r+1}\right)^2\right\} + \left\{\frac{r}{r+3} - \left(\frac{r}{r+3}\right)^2\right\} \\ &\quad + \left\{\frac{r}{r+5} - \left(\frac{r}{r+5}\right)^2\right\}. \end{aligned}$$

The solution corresponds to the one positive root of $U(\beta) = 0$, which can be written as $\phi = \arccos\{(45/23)\sqrt{3/23}\}$, $r = 2\sqrt{23/3}\cos(\phi/3) \approx 5.348721$, or $\hat{\beta} = \log(r) \approx 1.676858$.

Then

$LL(0) = -4.276666$	$LL(\hat{\beta}) = -3.358979$
$U(0) = 52/48$	$U(\hat{\beta}) = 0$
$\mathcal{I}(0) = 83/144$	$\mathcal{I}(\hat{\beta}) = 0.652077.$

The cumulative hazard now has a jump of size $1/(r+3) + 2/(r+5)$ at time 6. Efron [**?**] did not discuss estimation of the cumulative hazard, but it follows directly from the same argument as that used for the loglikelihood so we refer to it as the "Efron" estimate of the hazard. In S this hazard is the default whenever the Efron approximation for ties is used; the estimate is not available in SAS. For simple survival curves (i.e., the no-covariate case), the estimate is explored by Fleming and Harrington [**?**] as an alternative to the Kaplan–Meier.

The variance formula for the baseline hazard function is extended in the same way, and is the sum of (hazard increment)^2, treating a tied death as d separate hazard increments. In term 1 of the variance, the increment at time 6 is now $1/(r+3)^2 + 4/(r+5)^2$ rather than $2/(r+3)^2$. The increment to d at time 6 is $(r/(r+3)) * 1/(r+3) + (r/(r+5)) * 2/(r+5)$. (Numerically, the result of this computation is intermediate between the Nelson–Aalen variance and the Greenwood variance used in the Kaplan–Meier, which is an increment of

$$\frac{dN(t)}{[\sum Y_i(t) r_i(t)][\sum Y_i(t) r_i(t) - \sum dN_i(t) Y_i(t) r_i(t)]}.$$

The denominator for the Greenwood formula is the sum over those at risk, times that sum *without* the deaths. At time 6 this latter is $2/[(r+3)(3)]$.)

For $\beta = 0$, $x = 0$, let $v = \mathcal{I}^{-1} = 144/83$.

Time	Variance	
1	$1/36$	
	$+ v(1/12)^2$	$= 119/2988$
6	$(1/36 + 1/16 + 4/25)$	
	$+ v(1/12 + 1/16 + 1/18)^2$	$= 1996/6225$
9	$(1/36 + 1/16 + 4/25 + 1)$	
	$+ v(1/12 + 1/16 + 1/18 + 0)^2$	$= 8221/6225$

For $\beta = 1.676857$, $x = 0$.

Time	Variance	
1	0.00275667 + .00319386	= 0.0059505
2	0.05445330 + .0796212	= 0.134075
4	1.05445330 + .0796212	= 1.134075

Given the cumulative hazard, the martingale and score residuals follow directly using similar computations. Subject 3, for instance, experiences a total hazard of $1/(3r+3)$ at the first death time, $1/(r+3)$ at the "first" death on day 6, and $(1/2) * 2/(r+5)$ at the "second" death on day 6 — notice the case weight of 1/2 on the last term. Subjects 5 and 6 experience the full hazard of $1/(r+3)+2/(r+5)$ on day 6. The values of the martingale residuals are as follows.

Subject	$\widehat{M}(0)$	$\widehat{M}(\hat{\beta})$
1	5/6	.719171
2	–1/6	–.280829
3	5/12	–.438341
4	5/12	.731087
5	–3/4	–.365543
6	–3/4	–.365543

Let $a = r+1$, $b = r+3$, and $c = r+5$; then the score residuals are

Subject	Score	$L(0)$	$L(\hat{\beta})$
1	$2b/3a^2$	5/12	.113278
2	$-r/3a^2$	-1/12	–.044234
3	$1/3a^2 + a/2b^2 + b/2c^2$	55/144	–.102920
4	$r(1/3a^2 - a/2b^2 - b/2c^2$	–5/144	–.407840
5	$\frac{2r(177+282r+182r^2+50r^3+5r^4)}{3a^2b^2c^2}$	29/144	.220858
6	same	29/144	.220858

For subject 3, the score residual was computed as

$$\begin{aligned}\left(1-\tfrac{r}{r+1}\right)\left(0-\tfrac{1}{3r+3}\right) &+\left(1-\tfrac{r}{r+3}\right)\left(\tfrac{1}{2}-\tfrac{1}{r+3}\right)\\ &+\left(1-\tfrac{r}{r+5}\right)\left(\tfrac{1}{2}-\tfrac{1}{r+5}\right);\end{aligned}$$

the single death is counted as 1/2 event for each of the two day 6 events. Another equivalent approach is to actually form a second data set in which subjects 3 and 4 are each represented by two observations, one at time 6 and the other at time $6+\epsilon$, each with a case weight of 1/2. Then a computation using the Breslow approximation will give this score residual as the weighted sum of the score residuals for the two psuedo-observations.

The Schoenfeld residuals for the first and last events are identical to the Breslow estimates, that is, $1/(r+1)$ and 0, respectively. The residuals for time 6 are $1-c$ and $0-c$, where $c = (1/2)\{r/(r+3) + r/(r+5)\}$, the "average" $\bar{x}$ over the deaths.

It is quite possible to combine the Efron approximation for $\hat{\beta}$ along with the Breslow (or Nelson–Aalen) estimate of $\hat{\Lambda}$, and in fact this is the behavior used in some packages. That is, if the `ties=efron` option is chosen the formulas for LL, U, and $\mathcal{I}$ are those shown in this section, while the hazard and residuals all use the formulas of the prior section. Although this is not perfectly consistent the numerical effect on the residuals is minor, and it does not appear to affect their utility. S uses the calculations of this section by default.

The robust variance for a Cox model is defined as $D'D$ where the $n \times p$ dfbeta matrix D is based on the score residuals. Each row of D represents the infinitesimal jackknife, the derivative of $\hat{\beta}$ with respect to a change in the case weight for subject i. It is fairly each to check this using a direct derivative, $f(w_i + \epsilon) - f(w_i)/epsilon$ where f is the vector of coefficients from a fit of the Cox model with the chosen weight for subject i (w_i will be 1 for most data sets). This shows that the Efron/Breslow chimera is less accurate than the S code. However, I have not seen any example where the effect on either D or the robust variance $D'D$ was large enough to have practical consequences. Still, the numerical analyst in me prefers to avoid an inferior approximation.

1.3 Exact partial likelihood

At the tied death time the exact partial likelihood will have a single term. The numerator is a product of the risk scores of the subjects with an event, and the denominator is a sum of such products, where the sum is over all

possible choices of two subjects from the four who were at risk at the time. (If there were 10 tied deaths from a pool of 60 available, the sum would be over all $\binom{60}{10}$ subsets, a truly formidable computation!) In our case, three of the four subjects at risk at time 6 have a risk score of $\exp(0x) = 1$ and one a risk score of r, and the sum has six terms $\{r, r, r, 1, 1, 1\}$.

$$\begin{aligned} LL &= \{\beta - \log(3r+3)\} + \{\beta - \log(3r+3)\} + \{0 - 0\} \\ &= 2\{\beta - \log(3r+3)\}. \end{aligned}$$

$$\begin{aligned} U &= \left(1 - \frac{r}{r+1}\right) + \left(1 - \frac{r}{r+1}\right) + (0-0) \\ &= \frac{2}{r+1}. \end{aligned}$$

$$\mathcal{I} = \frac{2r}{(r+1)^2}.$$

The solution $U(\beta) = 0$ corresponds to $r = \infty$, with a loglikelihood that asymptotes to $-2\log(3)$. The Newton–Raphson iteration has increments of $(r+1)/r$, so $\hat{\beta} = 0$, 2, 3.1, 4.2, 5.2, and so on. A solution at $\hat{\beta} = 15$ is hardly different in likelihood from the true maximum, however, and most programs will stop iterating shortly after reaching this point. The information matrix, which measures the curvature of the likelihood function, rapidly goes to zero as β grows.

Both SAS and S use the Nelson–Aalen estimate of hazard after fitting an exact model, so the formulae of Table **??** apply. All residuals at $\hat{\beta} = 0$ are thus identical to those for a Breslow approximation. At $\hat{\beta} = \infty$ the martingale residuals are still well defined. Subjects 1 to 3, those with a covariate of 1, experience a hazard of $r/(3r+3) = 1/3$ at time 1. Subject 3 accumulates a hazard of 1/3 at time 1 and a further hazard of 2 at time 6. The remaining subjects are at an infinitely lower risk during days 1 to 6 and accumulate no hazard then, with subject 6 being credited with 1 unit of hazard at the last event. The residuals are thus $1 - 1/3 = 2/3$, $0 - 1/3$, $1 - 7/3 = -4/3$, $1 - 0$, 0, and 0, respectively, for the six subjects.

Values for the score and Schoenfeld residuals can be derived similarly as the limit as $r \to \infty$ of the formulae in Section **??**.

Time	Status	x	Number at Risk	$\bar{x}$	$d\hat{\Lambda}$
(1,2]	1	1	2	$r/(r+1)$	$1/(r+1)$
(2,3]	1	0	3	$r/(r+2)$	$1/(r+2)$
(5,6]	1	0	5	$3r/(3r+2)$	$1/(3r+2)$
(2,7]	1	1	4	$3r/(3r+1)$	$1/(3r+1)$
(1,8]	1	0	4	$3r/(3r+1)$	$1/(3r+1)$
(7,9]	1	1	5	$3r/(3r+2)$	$2/(3r+2)$
(3,9]	1	1			
(4,9]	0	1			
(8,14]	0	0	2	0	0
(8,17]	0	0	1	0	0

Table 2: Test data 2

2 Test data 2

This data set also has a single covariate, but in this case a (start, stop] style of input is employed. Table **??** shows the data sorted by the end time of the risk intervals. The columns for $\bar{x}$ and hazard are the values at the event times; events occur at the end of each interval for which status = 1.

2.1 Breslow approximation

For the Breslow approximation we have

$$\begin{aligned} LL &= \log\left(\frac{r}{r+1}\right) + \log\left(\frac{1}{r+2}\right) + \log\left(\frac{1}{3r+2}\right) + \\ &\quad \log\left(\frac{r}{3r+1}\right) + \log\left(\frac{1}{3r+1}\right) + 2\log\left(\frac{r}{3r+2}\right) \\ &= 4\beta - \log(r+1) - \log(r+3) - 3\log(3r+2) - 2\log(3r+1). \end{aligned}$$

$$\begin{aligned} U &= \left(1 - \frac{r}{r+1}\right) + \left(0 - \frac{r}{r+2}\right) + \left(0 - \frac{3r}{3r+2}\right) + \\ &\quad \left(1 - \frac{3r}{3r+1}\right) + \left(0 - \frac{3r}{3r+1}\right) + 2\left(1 - \frac{3r}{3r+2}\right) \\ &= 4 - \frac{63s^4 + 201s^3 + 184s^2 + 48s}{9s^4 + 36s^3 + 47s^2 + 24s + 4}. \end{aligned}$$

$$\mathcal{I} = \frac{r}{(r+1)^2} + \frac{2r}{(r+2)^2} + \frac{6r}{(3r+2)^2} + \frac{3r}{(3r+1)^2}$$

$$\frac{3r}{(3r+1)^2} + \frac{12r}{(3r+2)^2}.$$

The solution is at $U(\hat{\beta}) = 0$ or $r \approx .9189477$; $\hat{\beta} = \log(r) \approx -.084529$. Then

$$\begin{array}{ll} LL(0) = -9.392662 & LL(\hat{\beta}) = -9.387015 \\ U(0) = -2/15 & U(\hat{\beta}) = 0 \\ \mathcal{I}(0) = 2821/1800 & \mathcal{I}(\hat{\beta}) = 1.586935. \end{array}$$

The martingale residuals are (status–cumulative hazard) or $O - E = \delta_i - \int Y_i(s) r_i d\hat{\Lambda}(s)$. Let $\hat{\lambda}_1, \ldots, \hat{\lambda}_6$ be the six increments to the cumulative hazard listed in Table **??**. Then the cumulative hazards and martingale residuals for the subjects are as follows.

Subject	Λ_0	$\widehat{M}(0)$	$\widehat{M}(\hat{\beta})$
1	$r\hat{\lambda}_1$	1–30/60	0.521119
2	$\hat{\lambda}_2$	1–20/60	0.657411
3	$\hat{\lambda}_3$	1–12/60	0.789777
4	$r(\hat{\lambda}_2 + \hat{\lambda}_3 + \hat{\lambda}_4)$	1–47/60	0.247388
5	$\hat{\lambda}_1 + \hat{\lambda}_2 + \hat{\lambda}_3 + \hat{\lambda}_4 + \hat{\lambda}_5$	1–92/60	–0.606293
6	$r * (\hat{\lambda}_5 + \hat{\lambda}_6)$	1–39/60	0.369025
7	$r * (\hat{\lambda}_3 + \hat{\lambda}_4 + \hat{\lambda}_5 + \hat{\lambda}_6)$	1–66/60	–0.068766
8	$r * (\hat{\lambda}_3 + \hat{\lambda}_4 + \hat{\lambda}_5 + \hat{\lambda}_6)$	0–66/60	–1.068766
9	$\hat{\lambda}_6$	0–24/60	–0.420447
10	$\hat{\lambda}_6$	0–24/60	–0.420447

The score and Schoenfeld residuals can be laid out in a tabular fashion. Each entry in the table is the value of $\{x_i - \bar{x}(t_j)\} d\widehat{M}_i(t_j)$ for subject i and event time t_j. The row sums of the table are the score residuals for the subject; the column sums are the Schoenfeld residuals at each event time. Below is the table for $\beta = \log(2)$ ($r = 2$). This is a slightly more stringent test than the table for $\beta = 0$, since in this latter case a program could be missing a factor of $r = \exp(\beta) = 1$ and give the correct answer. However, the results are much more compact than those for $\hat{\beta}$, since the solutions are exact fractions.

	Event Time						Score
Id	2	3	6	7	8	9	Resid
1	$\frac{1}{9}$						$\frac{1}{9}$
2		$-\frac{3}{8}$					$-\frac{3}{8}$
3			$-\frac{21}{32}$				$-\frac{21}{32}$
4		$-\frac{1}{4}$	$-\frac{1}{16}$	$\frac{5}{49}$			$-\frac{165}{784}$
5	$\frac{2}{9}$	$\frac{1}{8}$	$\frac{3}{32}$	$\frac{6}{49}$	$-\frac{36}{49}$		$-\frac{2417}{14112}$
6					$-\frac{2}{49}$	$\frac{1}{8}$	$\frac{33}{392}$
7			$-\frac{1}{16}$	$-\frac{2}{49}$	$-\frac{2}{49}$	$\frac{1}{8}$	$-\frac{15}{784}$
8			$-\frac{1}{16}$	$-\frac{2}{49}$	$-\frac{2}{49}$	$-\frac{1}{8}$	$-\frac{211}{784}$
9						$\frac{3}{16}$	$\frac{3}{16}$
10						$\frac{3}{16}$	$\frac{3}{16}$
	$\frac{1}{3}$	$-\frac{1}{2}$	$-\frac{3}{4}$	$\frac{1}{7}$	$-\frac{6}{7}$	$\frac{1}{2}$	$-\frac{95}{84}$
	$\frac{1}{r+1}$	$\frac{-r}{r+2}$	$\frac{-3r}{r+2}$	$\frac{1}{3r+1}$	$\frac{3r}{3r+1}$	$\frac{4}{3r+2}$	

Both the Schoenfeld and score residuals sum to the score statistic $U(\beta)$. As discussed further above, programs will return two Schoenfeld residuals at time 7, one for each subject who had an event at that time.

2.2 Efron approximation

This example has only one tied death time, so only the term(s) for the event at time 9 change. The main quantities at that time point are as follows.

	Breslow	Efron
LL	$2\log\left(\frac{r}{3r+2}\right)$	$\log\left(\frac{r}{3r+2}\right) + \log\left(\frac{r}{2r+2}\right)$
U	$\frac{2}{3r+2}$	$\frac{1}{3r+2} + \frac{1}{2r+2}$
$\mathcal{I}$	$2\frac{6r}{(3r+2)^2}$	$\frac{6r}{(3r+2)^2} + \frac{4r}{(2r+2)^2}$
$d\hat{\Lambda}$	$\frac{2}{3r+2}$	$\frac{1}{3r+2} + \frac{1}{2r+2}$

Time	Status	X	Wt	$\bar{x}(t)$	$d\hat{\Lambda}_0(t)$
1	1	2	1	$(2r^2+11r)d\hat{\Lambda}_0 = \bar{x}_1$	$1/(r^2+11r+7)$
1	0	0	2		
2	1	1	3	$11r/(11r+5) = \bar{x}_2$	$10/(11r+5)$
2	1	1	4		
2	1	0	3		
2	0	1	2		
3	0	0	1		
4	1	1	2	$2r/(2r+1) = \bar{x}_3$	$2/(2r+1)$
5	0	0	1		

Table 3: Test data 3

3 Test data 3

This is very similar to test data 1, but with the addition of case weights. There are 9 observations, x is a 0/1/2 covariate, and weights range from 1 to 4. As before, let $r = \exp(\beta)$ be the risk score for a subject with $x = 1$. Table **??** shows the data set along with the mean and increment to the hazard at each point.

3.1 Breslow estimates

The likelihood is now a product of terms, one for each death, of the form

$$\left(\frac{e^{X_i\beta}}{\sum_j Y_j(t_i)w_j e^{X_j\beta}}\right)^{w_i}$$

leading to a log-likelihood very like equation **??**

$$l(\beta) = \sum_{i=1}^{n} \int_0^\infty \left[X_i(t)\beta - \log\left(\sum_j Y_j(t)w_j r_j(t)\right)\right] w_i dN_i(t) \qquad (1)$$

For integer weights, this gives the same results as would be obtained by replicating each observation the specified number of times, which was in fact one motivation for the definition. The definitions for the score vector U and information matrix $\mathcal{I}$ simply replace the mean and variance with weighted versions of the same. Let $l(\beta, w)$ be the logliklihood when all the observations are given a common case weight of w; it is easy to prove that $l(\beta, w) = wl(\beta, 1) - d\log(w)$ where d is the number of events. One consequence of this is that the log-likelihood can be positive when many of

the weights are < 1, which sometimes occurs in survey sampling applications. (This can be a big surprise the first time one encounters the situation.)

$$\begin{aligned} LL &= \{2\beta - \log(r^2 + 11r + 7)\} + 3\{\beta - \log(11r + 5)\} \\ &\quad +4\{\beta - \log(11r + 5)\} + 3\{0 - \log(11r + 5)\} \\ &\quad +2\{\beta - \log(2r + 1)\} \\ &= 11\beta - \log(r^2 + 11r + 7) - 10\log(11r + 5) - 2\log(2r + 1) \end{aligned}$$

$$\begin{aligned} U &= (2 - \bar{x}_1) + 3(0 - \bar{x}_2) + 4(1 - \bar{x}_2) + 3(1 - \bar{x}_2) + 2(1 - \bar{x}_3) \\ &= 11 - (\bar{x}_1 + 10\bar{x}_2 + 2\bar{x}_3) \\ I &= [(4r^2 + 11r)/(r^2 + 11r + 7) - \bar{x}_1^2] + 10(\bar{x}_2 - \bar{x}_2^2) + 2(\bar{x}_3 - \bar{x}_3^2) \end{aligned}$$

The solution corresponds to $U(\beta) = 0$, which is the solution point of the polynomial $66r^4 + 425r^3 - 771r^2 - 1257r - 385 = 0$, or $\hat{\beta} \approx \log(2.3621151) = 0.8595574$. Then

$$\begin{array}{ll} LL(0) = -32.86775 & LL(\hat{\beta}) = -32.02105 \\ U(0) = 2.107456 & U(\hat{\beta}) = 0 \\ \mathcal{I}(0) = 2.914212 & \mathcal{I}(\hat{\beta}) = 1.966563 \end{array}$$

When $\beta = 0$, the three unique values for $\bar{x}$ at $t = 1$, 2, and 4 are 13/19, 11/16 and 2/3, respectively, and the increments to the cumulative hazard are 1/19, $10/16 = 5/8$, and 2/3, see table **??**. The martingale and score residuals at $\beta = 0$ are $\hat{\beta}$ of

Id	Time	$M(0)$	$M(\hat{\beta})$
A	1	$1 - 1/19 = 18/19$	0.85531
B	1	$0 - 1/19 = -1/19$	-0.02593
C	2	$1 - (1/19 + 5/8) = 49/152$	0.17636
D	2	$1 - (1/19 + 5/8) = 49/152$	0.17636
E	2	$1 - (1/19 + 5/8) = 49/152$	0.65131
F	2	$0 - (1/19 + 5/8) = -103/152$	-0.82364
G	3	$0 - (1/19 + 5/8) = -103/152$	-0.34869
H	4	$1 - (1/19 + 5/8 + 2/3) = -157/456$	-0.64894
I	5	$0 - (1/19 + 5/8 + 2/3) = -613/456$	-0.69808

Score residuals at $\beta = 0$ are

Id	Time	Score
A	1	$(2-13/19)(1-1/19)$
B	1	$(0-13/19)(0-1/19)$
C	2	$(1-13/19)(0-1/19)+(1-11/16)(1-5/8)$
D	2	$(1-13/19)(0-1/19)+(1-11/16)(1-5/8)$
E	2	$(0-13/19)(0-1/19)+(0-11/16)(1-5/8)$
F	2	$(1-13/19)(0-1/19)+(1-11/16)(0-5/8)$
G	3	$(1-13/19)(0-1/19)+(0-11/16)(0-5/8)$
H	4	$(1-13/19)(0-1/19)+(1-11/16)(0-5/8)$ $+(1-2/3)(1-2/3)$
I	5	$(1-13/19)(0-1/19)+(1-11/16)(0-5/8)$ $+(0-2/3)(0-2/3)$

SAS returns the unweighted residuals as given above; it is the weighted sum of residuals that totals zero, $\sum w_i \widehat{M}_i = 0$, likewise for the score and Schoenfeld residuals evaluated at $\hat{\beta}$. S also returns unweighted residuals by default, with an option to return the weighted version. Whether the weighted or the unweighted form is more useful depends on the intended application, neither is more "correct" than the other. S does differ for the dfbeta residuals, for which the default is to return weighted values. For the third observation in this data set, for instance, the unweighted dfbeta is an approximation to the change in $\hat{\beta}$ that will occur if the case weight is changed from 2 to 3, corresponding to deletion of one of the three "subjects" that this observation represents, and the weighted form approximates a change in the case weight from 0 to 3, i.e., deletion of the entire observation.

The increments of the Nelson-Aalen estimate of the hazard are shown in the rightmost column of table **??**. The hazard estimate for a hypothetical subject with covariate $X^\dagger$ is $\Lambda_i(t) = \exp(X^\dagger\beta)\Lambda_0(t)$ and the survival estimate is $S_i(t) = \exp(-\Lambda_i(t))$. The two term of the variance, for $X^\dagger = 0$, are Term1 + $d'Vd$:

Time	Term 1
1	$1/(r^2+11r+7)^2$
2	$1/(r^2+11r+7)^2 + 10/(11r+5)^2$
4	$1/(r^2+11r+7)^2 + 10/(11r+5)^2 + 2/(2r+1)^2$

Time	d
1	$(2r^2+11r)/(r^2+11r+7)^2$
2	$(2r^2+11r)/(r^2+11r+7)^2 + 110r/(11r+5)^2$
4	$(2r^2+11r)/(r^2+11r+7)^2 + 110r/(11r+5)^2 + 4r/(2r+1)^2$

For $\beta = \log(2)$ and $X^\dagger = 0$, where $k \equiv$ the variance of $\hat{\beta} = 1/2.153895$ this reduces to

Time	Variance	
1	$1/1089$	$+ k(30/1089)^2$
2	$(1/1089 + 10/729)$	$+ k(30/1089 + 220/729)^2$
4	$(1/1089 + 10/729 + 2/25)$	$+ k(30/1089 + 220/729 + 8/25)^2$

giving numeric values of 0.0012706, 0.0649885, and 0.2903805, respectively.

3.2 Efron approximation

For the Efron approximation the combination of tied times and case weights can be approached in at least two ways. One is to treat the case weights as replication counts. There are then 10 tied deaths at time 2 in the data above, and the Efron approximation involves 10 different denominator terms. Let $a = 7r + 3$, the sum of risk scores for the 3 observations with an event at time 2 and $b = 4r + 2$, the sum of risk scores for the other subjects at risk at time 2. For the replication approach, the loglikelihood is

$$\begin{aligned} LL &= \{2\beta - \log(r^2 + 11r + 7)\} + \\ &\quad \{7\beta - \log(a + b) - \log(.9a + b) - \ldots - \log(.1a + b)\} + \\ &\quad \{2\beta - \log(2r + 1) - \log(r + 1)\}. \end{aligned}$$

A test program can be created by comparing results from the weighted data set (9 observations) to the unweighted replicated data set (19 observations). This is the approach taken by SAS `phreg` using the `freq` statement. It's advantage is that the appropriate result for all of the weighted computations is perfectly clear, the disadvantage is that only integer case weights are supported. (A second advantage is that I did not need to create another algebraic derivation for my test suite.)

A second approach, used in S, allows for non-integer weights. SAS also has weighted estimates, but I am not familiar with their algorithms. The data is considered to be 3 tied observations, and the log-likelihood at time 2 is the sum of 3 weighted terms. The first term of the three is one of

$$\begin{aligned} & 3[\beta - \log(a + b)] \\ & 4[\beta - \log(a + b)] \\ \text{or} \quad & 3[0 - \log(a + b)], \end{aligned}$$

depending on whether the event for observation C, D or E actually happened first (had we observed the time scale more exactly); the leading multiplier

of 3, 4 or 3 is the case weight. The second term is one of 6 possiblities

$$\begin{array}{rll} & 4[\beta - \log(4r+3+b)] & CDE \\ & 3[\beta - \log(4r+3+b)] & CED \\ & 3[0 - \log(3r+3+b)] & DCE \\ & 3[0 - \log(3r+3+b)] & DEC \\ & 3[\beta - \log(7r+0+b)] & ECD \\ \text{or} & 4[\beta - \log(7r+0+b)] & EDC \end{array}$$

The first choice corresponds to an event order of observation C then D (subject D has the event, with D and E still at risk), etc. For a weighted Efron approximation first replace each term by its average, just as in the unweighted case. The first terms ends up as $(7/3)\beta - (10/3)\log(a+b)$, the second as $(7/3)\beta - 20/6\log(2a/3+b)$, and the third as $(7/3)\beta - 10/3\log(a/3+b)$. This replaces the interior of the log function with its average, and the multiplier of the log with the average weight of 10/3.

The final log-likelihood and score statistic are

$$\begin{aligned} LL &= \{2\beta - \log(r^2+11r+7)\} \\ &\quad +\{7\beta - (10/3)[\log(a+b) + \log(2a/3+b) + \log(a/3+b)]\} \\ &\quad +2\{\beta - \log(2r+1)\} \end{aligned}$$

$$\begin{aligned} U &= (2-\bar{x}_1) + 2(1-\bar{x}_3) \\ &\quad +7 - (10/3)[\bar{x}_2 + 26r/(26r+12) + 19r/(19r+9)] \\ &= 11 - (\bar{x}_1 + (10/3)(\bar{x}_2 + \bar{x}_{2b} + \bar{x}_{2c}) + 2\bar{x}_3) \end{aligned}$$

$$\begin{aligned} I &= [(4s^2+11s)/(s^2+11s+7) - \bar{x}_1^2] \\ &\quad +(10/3)[(\bar{x}_2 - \bar{x}_2^2) + (\bar{x}_{2b} - \bar{x}_{2b}^2) + (\bar{x}_{2b} - \bar{x}_{2b}^2) \\ &\quad +2(\bar{x}_3 - \bar{x}_3^2) \end{aligned}$$

The solution is at $\beta = .87260425$, and

$$\begin{array}{ll} LL(0) = -30.29218 & LL(\hat{\beta}) = -29.41678 \\ U(0) = 2.148183 & U(\hat{\beta}) = 0 \\ \mathcal{I}(0) = 2.929182 & \mathcal{I}(\hat{\beta}) = 1.969447\,. \end{array}$$

The hazard increment and mean at times 1 and 4 are identical to those for the Breslow approximation, as shown in table **??**. At time 2, the number

at risk for the first, second and third portions of the hazard increment are $n_1 = 11r + 5$, $n_2 = (2/3)(7r + 3) + 4r + 2 = (26r + 12)/3$, and $n_3 = (1/3)(7r + 3) + 4r + 2 = (19r + 9)/3$. Subjects F–I experience the full hazard at time 2 of $(10/3)(1/n_1 + 1/n_2 + 1/n_3)$, subjects B–D experience $(10/3)(1/n_1 + 2/3n_2 + 1/3n_3)$. Thus, at $\beta = 0$ the martingale residuals are

Id	Time	$\widehat{M}(0)$	
A	1	1 - 1/19	= 18/19
B	1	0 - 1/19	= -1/19
C	2	1 - (1/19 + 10/48 + 20/114 + 10/84)	=473/1064
D	2	1 - (1/19 + 10/48 + 20/114 + 10/84)	=473/1064
E	2	1 - (1/19 + 10/48 + 20/114 + 10/84)	=473/1064
F	2	0 - (1/19 + 10/48 + 10/38 + 10/28)	=-2813/3192
G	3	0 - (1/19 + 10/48 + 10/38 + 10/28)	=-2813/3192
H	4	1 - (1/19 + 10/48 + 10/38 + 10/28 + 2/3)	=-1749/3192
I	5	0 - (1/19 + 10/48 + 10/38 + 10/28 + 2/3)	=-4941/3192

The hazard estimate for a hypothetical subject with covariate $X^\dagger$ is $\Lambda_i(t) = \exp(X^\dagger\beta)\Lambda_0(t)$, Λ_0 has increments of $1/(r^2+11r+7$, $(10/3)(1/n_1 + 1/n_2 + 1/n_3)$ and $2/(2r + 1)$. This increment at time 2 is a little larger than the Breslow jump of $10/d1$. The first term of the variance will have an increment of $[\exp((X^\dagger\beta)[]^2(10/3)(1/n_1^2 + 1/n_2^2 + 1/n_3^2)$ at time 2. The increment to the cumulative distance from the center d will be

$$\begin{aligned} & \left[X^\dagger - \frac{11r}{11r+5}\right]\frac{10}{3n_1} \\ + & \left[X^\dagger - \frac{(2/3)7r+4r}{n2}\right](10/3)(1/n_2) \\ + & \left[X^\dagger - \frac{(1/3)7r+4r}{n2}\right](10/3)(1/n_3) \end{aligned}$$

For $X^\dagger = 1$ and $\beta = \pi/3$ we get cumulative hazard and variance below. We have $r = e^\pi/3$,

Λ		Variance
$e/(r^2 + 11r + 7)$	0.03272832	$e^2/(r^2 + 11r + 7)^2$

4 Test data 4

This is an extension of test data set 3, but with 3 covariates. Let $r_i = \exp(X_\beta)$ be the risk score for each subject, β unspecified. Table **??** shows the data set along with the mean and increment to the hazard at each point.

At $\beta = 0$ we have $r = 1$ and

Id	Time	Status	x_1	x_2	x_3	Wt	Denominator	$\bar{x}_2$
1	10	1	0	2	5	1	$d_1 = r_1 + 2r_2 + d_2 + d_3$	$(2r_1 + 3r_3 + 4r_4 + 2r_6 + 2r_8)/d_1$
2	10	0	0	0	2	2		
3	20	1	1	1	3	3	$d_2 = 3r_3 + 4r_4 + 3r_5 +$	$(3r_3 + 4r_4 + 2r_6 + 2r_8)/d_2$
4	20	1	1	1	6	4	$2r_6 + r_7 + d_3$	
5	20	1	0	0	4	3		
6	20	0	0	1	3	2		
7	30	0	1	0	1	1		
8	40	1	1	1	3	2	$d_3 = 2r_8 + r_9$	$2r_4/d_3$
9	50	0	1	0	1	1		

Table 4: Test data 4