Relative-risk models

26 February, 2020

Cox proportional hazards model (relative-risk regression)

Semiparametric model $h_i(t) = h\left(t \mid \mathbf{x}_i\right) = h_0(t) r\left(\beta, \mathbf{x}_i(t); t\right)$.

$\mathbf{x}_i$ = vector of individual covariates
β = vector of parameters

Usually assume *r* has no direct dependence on *t*.

Cox model: Generalised linear model with log link function, so $r(\beta, \mathbf{x}) = e^{\beta^\top \mathbf{x}} = e^{\sum \beta_j x_j}$.

Alternative: Excess relative risk model $r(\beta, \mathbf{x}) = \prod_{j=1}^{p} \left(1 + \beta_j x_j\right)$

Linear relative risk $r(\beta, \mathbf{x}) = 1 + \sum_{j=1}^{p} \beta_j x_j$.

We will consider only the Cox model in detail.

Important points:

1. Covariates should be centred.
2. Categorical covariates must be represented as binary variables.

Partial likelihood

We are given the data

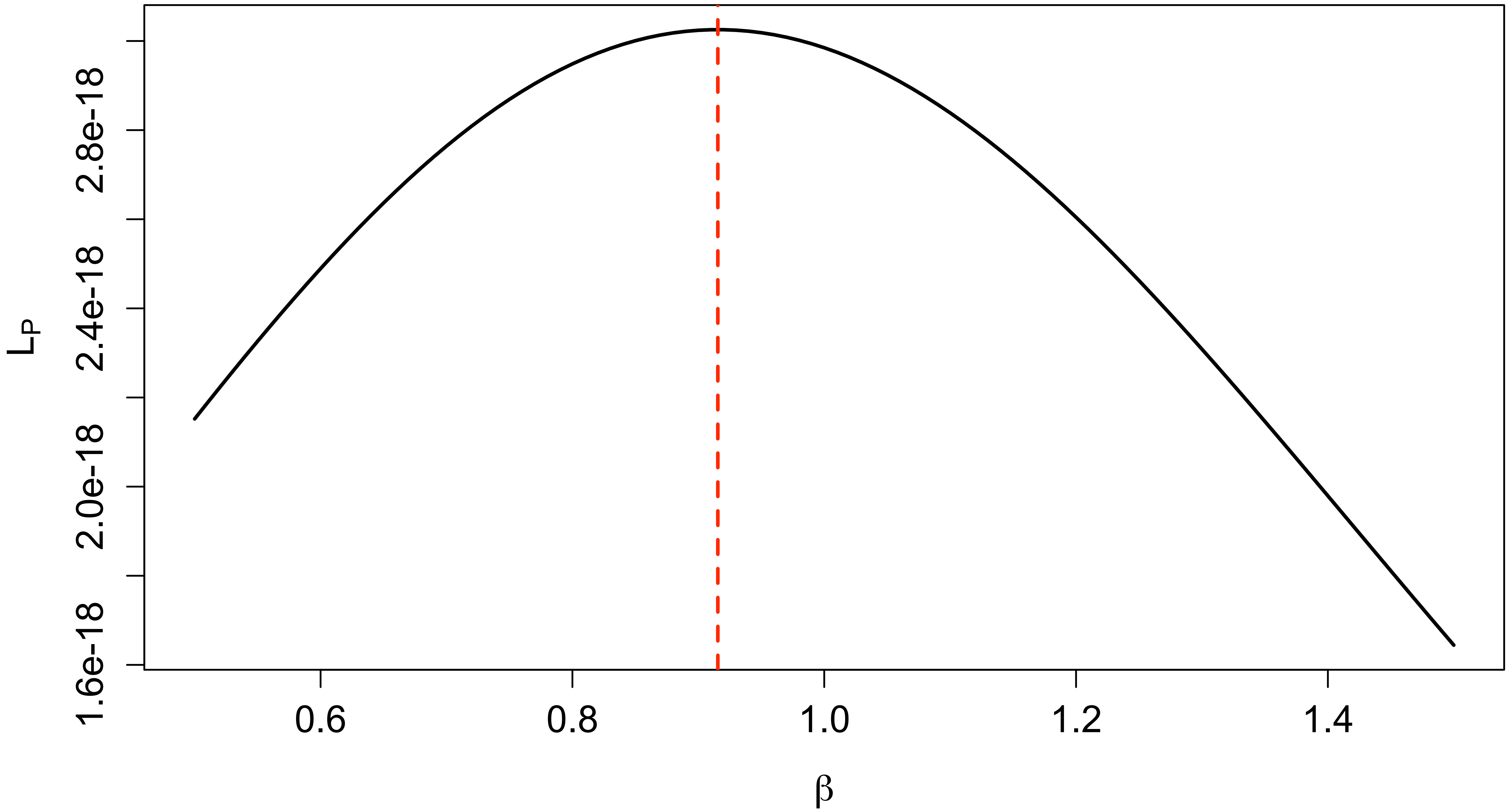
- A list of event times $t_1 < t_2 < \cdots$. (We are assuming no ties, for the moment.)
- The identity i_j of the individual whose event is at time t_j .
- The values of all individuals' covariates (at times t_j, if they are varying).
- The risk sets $\mathscr{R}_j = \{i \,:\, T_i \geq t_j\}$, the set of individuals who are at risk at time t_j .

Partial likelihood is the conditional probability of the **choice** of subject who has an event.

$$\pi\big(i \mid t\big) := \frac{h_i(t)}{h(t)} = \frac{r(\beta, \mathbf{x}_i(t); t)\mathbf{1}_{\{i \in \mathscr{R}(t)\}}}{\sum_{j \in \mathscr{R}(t)} r(\beta, \mathbf{x}_j(t); t)},$$

$$L_P(\beta) = \prod_{t_j} \pi\big(i_j \mid t_j\big) = \prod_{t_j} \frac{r(\beta, \mathbf{x}_{i_j}(t_j); t_j)}{\sum_{l \in \mathscr{R}_j} r(\beta, \mathbf{x}_l(t_j); t_j)} .$$

Theorem: Let $\hat{\beta}$ maximise L_P. Then $\hat{\beta}$ is a consistent estimator of the true parameter β_0, and $\sqrt{n}(\hat{\beta} - \beta_0)$ is asymptotically normal with variance consistently approximated by $\mathscr{J}_P^{-1}(\hat{\beta})$, the matrix of second derivatives of the log partial likelihood.

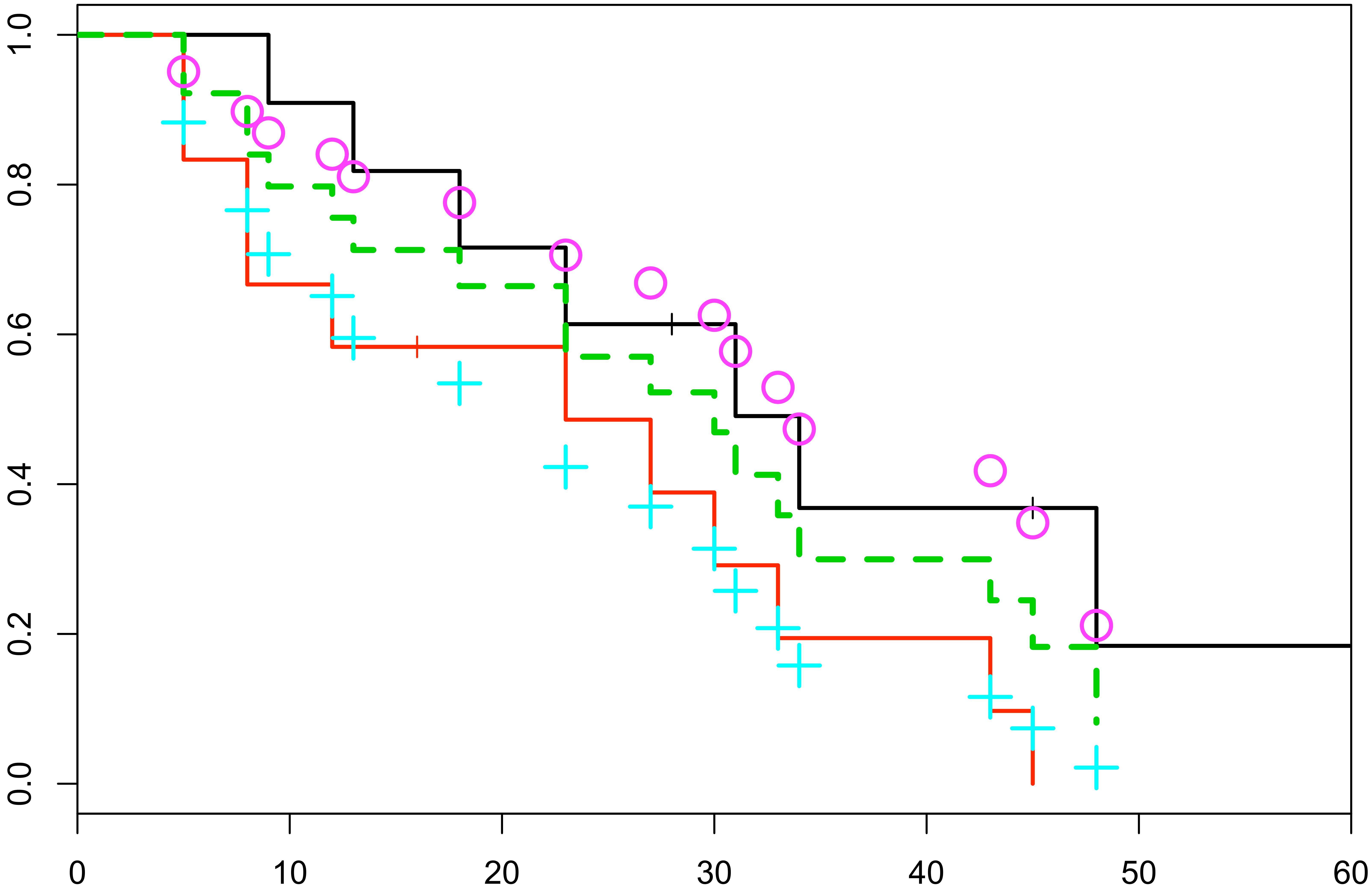


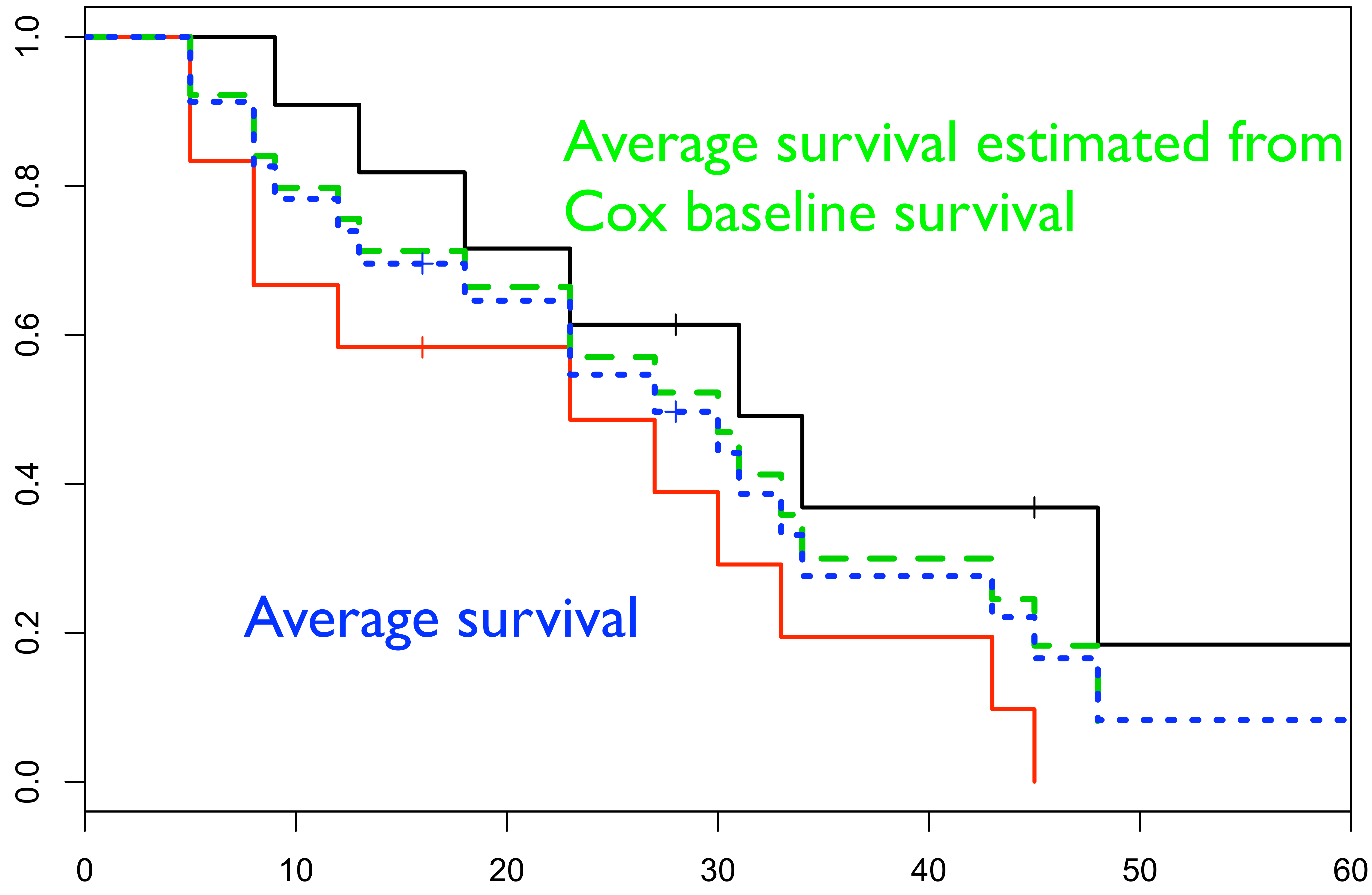
$$L_P(\beta) = \left(\frac{e^{2\beta}}{(12e^{\beta}+11)(11e^{\beta}+11)}\right)\left(\frac{e^{2\beta}}{(10e^{\beta}+11)(9e^{\beta}+11)}\right)$$
$$\times\left(\frac{1}{8e^{\beta}+11}\right)\left(\frac{e^{\beta}}{8e^{\beta}+10}\right)\left(\frac{1}{7e^{\beta}+10}\right)\left(\frac{1}{6e^{\beta}+8}\right)$$
$$\times\left(\frac{e^{\beta}\cdot 1}{(6e^{\beta}+7)(5.5e^{\beta}+6.5)}\right)\left(\frac{e^{\beta}}{5e^{\beta}+6}\right)\left(\frac{e^{\beta}}{4e^{\beta}+5}\right)$$
$$\times\left(\frac{1}{3e^{\beta}+5}\right)\left(\frac{e^{\beta}}{3e^{\beta}+4}\right)\left(\frac{1}{2e^{\beta}+4}\right)\left(\frac{e^{\beta}}{2e^{\beta}+3}\right)\left(\frac{e^{\beta}}{e^{\beta}+3}\right)\left(\frac{1}{2}\right)$$

```
coxph(formula = Surv(time, status) ~ x, data = aml)
                  coef exp(coef) se(coef)    z     p
xNonmaintained 0.916       2.5     0.512 1.79 0.074

Likelihood ratio test=3.38  on 1 df, p=0.0658  n= 23
```

```
coxph(formula = Surv(time, status) ~ x, data = aml,
                  method = "breslow")
                  coef exp(coef) se(coef)    z     p
xNonmaintained 0.904      2.47     0.512 1.77 0.078

Likelihood ratio test=3.3  on 1 df, p=0.0694  n= 23
```

estimate β=0.916




Average survival estimated from
Cox baseline survival
Average survival
1.0
0.8
0.6
0.4
0.2
0.0
0
10
20
30
40
50
60

Significance tests

Estimate variance for the parameters from observed Fisher partial information.

Single parameter: Test H_0:β_q=0 using asymptotically normal test statistic $\hat{\beta}_q\sqrt{\mathcal{I}_P(\hat{\beta})_{qq}}$.

As with the MLE, there are three (asymptotically equivalent) test statistics conventionally used to test the null hypothesis $\beta = \beta_0$:

- **Wald statistic**: $\xi^2_W := (\hat{\beta} - \beta_0)^T \mathcal{J}(\beta_0)(\hat{\beta} - \beta_0)$;
- **Score statistic**: $\xi^2_{SC} = U(\beta_0)^T \mathcal{J}(\beta_0) U(\beta_0)$;
- **Likelihood ratio statistic**: $\xi^2_{LR} = 2\big[\ell_P(\hat{\beta}) - \ell_P(\beta_0)\big],$ where $\ell_P := \log L_P$.

Under the null hypothesis these are all asymptotically chi-squared distributed with *p* degrees of freedom.

Baseline hazard

Nelson—Aalen: $\hat{H}(t) = \sum_{t_j \le t} \frac{d_j}{n_j}$. Here $h_j = \frac{d_j}{n_j}$ is the fraction of the population at risk at time t_j that has an event at time t_j .

In the relative risk setting if a single individual has an event at t_j then the *fraction of risk* that has an event at time t_j is $\frac{r(\beta, x_i(t_j))}{\sum_{i' \in \mathcal{R}_j} r(\beta, x_{i'}(t_j))} = \widehat{h_i}(t_j) = r(\beta, x_i(t_j))\,\widehat{h_0}(t_j)$.

Breslow's estimator: $$\widehat{H_0}(t) = \sum_{t_j \le t} \frac{1}{\sum_{i \in \mathcal{R}_j} r(\beta, x_i(t_j))}$$

$$= \sum_{t_j \le t} \frac{1}{\sum_{i \in \mathcal{R}_j} e^{\beta \cdot x_i(t_j)}}$$ in the case of Cox regression.

Individual risk

An individual with constant covariate x may be estimated to have a cumulative hazard

$$\widehat{H}(t\,|\,\mathbf{x}) = r(\hat{\beta}, \mathbf{x})\,\widehat{H_0}(t)$$

For time-varying covariates $H(t\,|\,\mathbf{x}) = \int_0^t r(\beta, \mathbf{x}(u))h_0(u)du$, which we approximate by

$$\hat{H}(t\,|\,\mathbf{x}) = \sum_{t_j \le t} r(\hat{\beta}, \mathbf{x}(t_j))\,\widehat{h_0}(t_j) = \sum_{t_j \le t} \frac{r(\hat{\beta}, \mathbf{x}(t_j))}{\sum_{i \in \mathcal{R}_j} r(\hat{\beta}, \mathbf{x}_i(t_j))}\,.$$

For Cox regression this becomes

$$\hat{H}(t\,|\,\mathbf{x}) = \sum_{t_j \le t} \mathrm{e}^{\hat{\beta}\cdot\mathbf{x}(t_j)}\,\widehat{h_0}(t_j) = \sum_{t_j \le t} \frac{\mathrm{e}^{\hat{\beta}\cdot\mathbf{x}(t_j)}}{\sum_{i \in \mathcal{R}_j} \mathrm{e}^{\hat{\beta}\cdot\mathbf{x}_i(t_j)}}\,.$$

Tied data

Suppose at time t_j there are 5 individuals at risk, with relative risks r_1, r_2, r_3, r_4, r_5.
Individuals 1 and 2 have events at time t_j.

The contribution to the partial likelihood is

$$\frac{r_1}{r_1+r_2+r_3+r_4+r_5}\cdot\frac{r_2}{r_2+r_3+r_4+r_5}+\frac{r_2}{r_1+r_2+r_3+r_4+r_5}\cdot\frac{r_1}{r_1+r_3+r_4+r_5}.$$

In general, if d_j individuals have events, there are $d_j!$ terms.

Efron's alternative:

$$\frac{r_1 r_2}{(r_1+r_2+r_3+r_4+r_5)\left(\frac{1}{2}(r_1+r_2)+r_3+r_4+r_5\right)}.$$

In other words,

$$\ell_P(\beta)=\sum_{t_j}\sum_{i\in\mathcal{D}_j}\log r(\beta,\mathbf{x}_i(t_j))-\sum_{k=0}^{d_j-1}\log\left(\sum_{i\in\mathcal{R}_j} r(\beta,\mathbf{x}_i(t_j)-\frac{k}{d_i}\sum_{i\in\mathcal{D}_j} r(\beta,\mathbf{x}_i)\right).$$

Applying the same principle to baseline hazard estimate,

$$\widehat{H_0}(t)=\sum_{t_j\le t}\sum_{k=0}^{d_j-1}\left(\sum_{i\in\mathcal{R}_j} r\big(\beta,\mathbf{x}_i(t_j)\big)-\frac{k}{d_i}\sum_{i\in\mathcal{D}_j} r\big(\beta,\mathbf{x}_i(t_j)\big)\right)^{-1}.$$

Breslow's alternative: Ignore the decrements to "total hazard at risk"

$$\ell_P^{Breslow}(\beta)=\sum_{t_j}\sum_{i\in\mathcal{D}_j}\log r(\beta,\mathbf{x}_i(t_j))-d_j\log\sum_{i\in\mathcal{R}_j} r\big(\beta,\mathbf{x}_i(t_j)\big).$$