Empirical likelihood

Definition

The empirical likelihood with right censoring is

$$L(\rho) = \prod_i \rho(T_i)^{e_i} S(T_i)^{1-e_i}.$$

Just as in the uncensored setup the empirical likelihood will give the total likelihood ordering on the set of distributions with point masses in the observations.

Discrete hazards

Define

$$\lambda_i = \frac{\rho(T_i)}{S(T_i-)} = P(T^* = T_i \mid T^* \geq T_i).$$

Lemma

For a discrete probability measure with point masses in $T_1, \ldots, T_n$ and corresponding hazards λ_i the survival function is

$$S(t) = \prod_{i: T_i \leq t} (1 - \lambda_i).$$

NPMLE for censored observations

Theorem

The NPMLE of the hazards – the hazards for the probability measure that maximize L – is given by

$$\hat{\lambda}_i = \frac{e_i}{Y_i}$$

where $Y_i = Y(T_i) = \sum_j 1(T_i \leq T_j)$ *is the number of individuals at risk at time* T_i*.*

The Nelson-Aalen estimator of the cumulative (discrete) hazard is

$$\hat{\Lambda}(t) = \sum_{i: T_i \leq t} \frac{e_i}{Y_i}.$$

And the NPMLE of the survival function is the Kaplan-Meier estimator.

Poisson number of events

The likelihood can be written as

$$L = \prod_i \lambda_i^{e_i}(1-\lambda_i)^{1-e_i} \prod_{j:T_j<T_i}(1-\lambda_j).$$

The ith factor is the probability that individual i does not die at any time T_j for $T_j < T_i$ (factorized as conditional probabilities) and then either survive until T_i and becomes censored, or dies at time T_i.

Suppose that the "individual" can experience more than one event simultaneous at the event time, and that the number of events in T_j conditionally on no events before time T_j is Poisson with mean λ_j.

The conditional probability of no events in T_j given no events before T_j is $e^{-\lambda_j}$.

Poisson empirical likelihood

The probability of no events before T_i is therefore

$$\prod_{j:T_j<T_i} e^{-\lambda_j} = e^{-\Lambda(T_i)+\lambda_i},$$

the conditional probability of no events at time T_i is $e^{-\lambda_i}$, and the conditional probability of one event at time T_i is $\lambda_i e^{-\lambda_i}$.

Definition

The Poisson empirical likelihood is

$$L^*(\Lambda) = \prod_i \lambda_i^{e_i} e^{-\Lambda_i}$$

where $\Lambda_i = \Lambda(T_i)$.

Proportional hazards model

Assume that the ith individual has hazard function

$$\lambda_i(t) = w_i \lambda_0(t),$$

for $w_i > 0$ a weight. The cumulative hazard function is $\Lambda_i(t) = w_i \Lambda_0(t)$.

Theorem

With $\Lambda_i(t) = w_i \Lambda_0(t)$, $W(t) = \sum_{j : t \leq T_j} w_j$ and $W_i = W(T_i)$ it holds that

$$L = \prod_{i : e_i = 1} \frac{w_i}{W_i} \left(\prod_i (W_i \lambda_0(T_i))^{e_i} \right) e^{-\int_0^\infty W(t) \lambda_0(t) \mathrm{d}t}.$$

Partial likelihood

Definition

Cox's partial likelihood is

$$L_{\text{par}} = \prod_{i:e_i=1} \frac{w_i}{W_i}.$$

Parameters that enter into the individual weights are estimated nonparametrically by mazimizing the partial likelihood.

The other factor in the full likelihood could potentially be informative, if we have a parametric model for the baseline λ_0, but becomes uninformative if we leave λ_0 unspecified.

Profiling the Poisson empirical likelihood

With the discrete baseline hazards parametrized as $\lambda_{0,j} = e^{h_j}$ the Poisson empirical log-likelihood becomes

$$\ell^* = \sum_i e_i h_i + e_i \log(w_i) - e^{h_i} W_i.$$

Lemma

The maximizer of ℓ^* over $h_1, \ldots, h_n$ is given by

$$e^{h_i} = \frac{e_i}{W_i}$$

for $i = 1, \ldots, n$.

The profile log-likelihood becomes

$$\ell^* = \underbrace{\sum_{i: e_i = 1} \log\left(\frac{w_i}{W_i}\right)}_{\text{Cox's partial log-likelihood}} - N, \qquad N = \sum_i e_i.$$

Estimating parameters in a regression model

With

$$\eta_i = X_i^T \beta$$

the linear predictor the ith weight is usually specified as $w_i = e^{\eta_i}$.

The partial log-likelihood becomes

$$\sum_{i=1}^{n} \log \left(\frac{e^{\eta_i}}{\sum_{j: T_i \leq T_j} e^{\eta_j}} \right) = \sum_{i=1}^{n} X_i^T \beta - \log \left(\sum_{j: T_i \leq T_j} e^{X_j^T \beta} \right),$$

which is then maximized to find estimates of β.

Standard likelihood arguments regarding quadratic approximations and approximate distributional results apply when using the partial likelihood.

Cox-Snell residuals ...

... are not really residuals. If T has continuous survival function S then

$$S(T) \sim \mathrm{unif}([0,1]).$$

Hence

$$\Lambda(T) \sim \mathrm{Exp}(1).$$

With $(T_1, e_1), \ldots, (T_n, e_n)$ right censored (independent censoring) survival times,

$$(\Lambda_1(T_1), e_1), \ldots, (\Lambda_n(T_n), e_n)$$

are independently right censored exponentially distributed.

Computing and plotting Cox-Snell

Fit a nonparametric cumulative hazard function to the Cox-Snell residuals

$$\hat{\Lambda}_i(T_i).$$

Plot the fit against time – should be a line with slope 1 (the cumulative hazard function for the exponential function).

Other residuals

With $T = \min\{T^*, C\}$, T^* and C having continuously differentiable survival functions S and H,

$$E\Lambda(T) = P(T^* \leq C).$$

by partial integration (computation below).

Hence with $e = 1(T^* \leq C)$,

$$E(e - \Lambda(T)) = 0.$$

$$\begin{aligned} E\Lambda(T) &= -\int_0^\infty \Lambda(t)(SH)'(t)\mathrm{d}t = \int_0^\infty \Lambda'(t)S(t)H(t)\mathrm{d}t \\ &= \int_0^\infty f(t)H(t)\mathrm{d}t = P(T^* < C). \end{aligned}$$

Martingale residuals

The martingale residuals are

$$e_i - \hat{\Lambda}_i(T_i)$$

They can be used much as regression residuals to judge overall fit as well as fit of the individual predictors. They have a left skewed distribution on $(-\infty, 1)$.

Poisson log-likelihood

The Poisson nonparametric log-likelihood as well as the parametric loglikelihood with a continuous baseline can be written as

$$\ell^* = \sum_{i=1}^{n} \underbrace{e_i \log(w_i \Lambda_{0,i}) - w_i \Lambda_{0,i}}_{\text{Poisson log-likelihood term}} + \sum_{i=1}^{n} e_i \log\left(\frac{\lambda_{0,i}}{\Lambda_{0,i}}\right)$$

The first term is identical to the Poisson log-likelihood, and the second term does not depend on parameters that enter into the weights. This inspires the definition of deviance residuals in terms of the Poisson deviances

$$d(e_i, \hat{\Lambda}_i) = 2\left(e_i \log(e_i/\hat{\Lambda}_i) - e_i + \hat{\Lambda}_i\right),$$

where $\hat{\Lambda}_i = \hat{w}_i \hat{\Lambda}_{0,i}$.

Deviance residuals

The result is the deviance residuals

$$\mathrm{sign}(e_i - \hat{\Lambda}_i)\sqrt{d(e_i, \hat{\Lambda}_i)} = \begin{cases} -\sqrt{2\hat{\Lambda}_i} & \text{if } e_i = 0 \\ \mathrm{sign}(1 - \hat{\Lambda}_i)\sqrt{2(\hat{\Lambda}_i - 1 - \log \hat{\Lambda}_i)} & \text{if } e_i = 1 \end{cases}$$

Observe that the deviance residuals can be expressed as

$$\mathrm{sign}(\hat{r}_i)\sqrt{-2(\hat{r}_i + e_i \log(e_i - \hat{r}_i))}$$

with $\hat{r}_i = e_i - \hat{\Lambda}_i$ the martingale residuals. Compared to the martingale residuals, the distribution of the deviance residuals is less skewed.