2 Definitions

In what follows, let $y_1,...,y_n$ be independent observations and let $x_1,...,x_n$ be associated explanatory variables , $x_i \in \mathbf{R}^p$. Assume that the estimate of interest $\hat{\theta}_n$ is defined by an estimating equation of the form

$$\sum_{i=1}^{n} h_{in}(y_i, x_i, \hat{\theta}_n) = 0 . \tag{1}$$

In many cases we can find a sequence of matrices $A_n = A_n(y_1,...,y_n, x_1,...,x_n,\theta) = A_n(\theta)$ such that

$$\sqrt{n}\, A_n(\theta)\left[\frac{1}{n}\sum_{i=1}^{n} h_{in}(y_i, x_i, \theta)\right] \xrightarrow[n\to\infty]{D} N(0,\Sigma), \tag{2}$$

where Σ is the asymptotic covariance matrix of the sequence $\hat{\theta}_n$, i.e.:

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow[n\to\infty]{D} N(0,\Sigma).$$

In the rest of this section Salibian-Barrera and Zamar have discussed three examples that illustrate the possible choices of h_{in} , i =1,…,n and A_n above. The three examples are robust location estimates, quasi-likelihood estimates and robust linear regression.

Example 2 - Quasi-likelihood estimates: Consider n independent random variables $Y_1,...,Y_n$ such that

$$E(Y_i \backslash X_i) = \mu_i = g(\beta^T X_i) \quad 1 \le i \le n,$$

where $X_1,...,X_n$ are p-dimensional vectors of explanatory variables and the link function g is known. Furthermore assume that the variance of Y_i is proportional to a known function υ of its mean:

$$V(Y_i \backslash X_i) = \sigma^2 \upsilon(\mu_i).$$

The quasi-likelihood equations [see Wedderburn (1974) and McCullagh (1983)] are

$$\sum_{i=1}^{n} \frac{\partial \mu_i}{\partial \beta}\left(\frac{y_i - \mu_i}{\sigma^2 \upsilon(\mu_i)}\right) = 0,$$

A Taylor expansion around β gives

$$\hat{\beta}_n - \beta = A_n^{-1}(\beta)\sum_{i=1}^{n}\frac{\partial\mu_i}{\partial\beta}\left(\frac{y_i - \mu_i}{\upsilon(\mu_i)}\right) + O_p(1/n), \qquad (9)$$

where

$$A_n(\beta) = \sum_{i=1}^{n}\frac{\partial\mu_i}{\partial\beta}\frac{1}{\upsilon(\mu_i)}\frac{\partial\mu_i^T}{\partial\beta}. \qquad (10)$$

Hence, to do inference for β we can use an asymptotically normal approximation. For example [see McCullagh (1983)] we have

$$\sqrt{n}(\hat{\beta}_n - \beta) = N_p(0, n\sigma^2\hat{A}_n^{-1}) + O_p(1/\sqrt{n}),$$

where $\hat{A}_n = A_n(\hat{\beta}_n)$ is the empirical estimate for A_n in (10). Moreover, cov ($\hat{\beta}_n$) ~ $\hat{\sigma}_n^2\hat{A}_n^{-1}$ where

$$\hat{\sigma}_n^2 = \frac{1}{n-p}\sum_{i=1}^{n}\frac{(y_i - \hat{\mu}_i)^2}{\upsilon(\hat{\mu}_i)} \ .$$

When the variance of the response variable is misspecified in the quasi-likelihood equations (i.e. $V(Y_i \backslash X_i) \neq \sigma^2\upsilon(\mu_i)$) we can still obtain a consistent estimate of cov ($\hat{\beta}_n$)

$$\text{cov}\ (\hat{\beta}_n) \sim \hat{A}_n^{-1}\hat{\Sigma}_n\hat{A}_n^{-1}\ , \qquad (11)$$

where

$$\hat{\Sigma}_n = \sum_{i=1}^{n}\frac{\partial\mu_i}{\partial\beta}\frac{(y_i - \hat{\mu}_i)^2}{\upsilon(\hat{\mu}_i)^2}\frac{\partial\mu_i^T}{\partial\beta}.$$

This is known as the "sandwich estimate" or the "robust covariance estimate" [see, for example, Eicker (1963); Huber (1967); White (1982); Gourieroux et al. (1984); Liang and Zeger (1986)]. Not surprisingly, this estimate is notably inefficient [see… Breslow (1990); … Carroll and Kauermann (2001)] . Hence, it is of interest to obtain an alternative inference method for these models that will be more efficient at the parametric model. Equation (9) clearly suggests the choices

$$h_{in}(y_i, x_i, \beta) = h(y_i, x_i, \beta) = \frac{\partial\mu_i}{\partial\beta}\frac{(y_i - \mu_i)}{\upsilon(\mu_i)} = \frac{\partial\mu_i}{\partial\beta}\frac{(y_i - g(\beta^T x_i)}{\upsilon(\beta^T x_i)},$$

and

$$A_n(\beta) = \sum_{i=1}^{n} \frac{\partial \mu_i}{\partial \beta} \frac{1}{\upsilon(\mu_i)} \frac{\partial \mu_i^T}{\partial \beta}.$$

3 Fast Bootstrap

To obtain a fast bootstrap method to approximate the distribution of $\hat{\theta}_n$ we bootstrap the left hand side of (2) replacing θ with $\hat{\theta}_n$. Recall that the sequence of matrices $A_n(\theta)$ can sometimes be written as $A_n(\theta)=B_n(\theta)C_n(\theta)$ where $B_n(\theta)$ and $C_n(\theta)$ also depend on the sample.

Consider the following two approaches:

- **Approach A:** The first approach is to bootstrap equation (2) maintaining "the denominator" fixed in other words, if $x_1^*,...,x_n^*$ is a bootstrap sample, then we recalculate the above equation

$$\sqrt{n} B_n(\hat{\theta}_n) C_n^*(\hat{\theta}_n) \left[\frac{1}{n} \sum_{i=1}^{n} h_{in}(x_i^*, \hat{\theta}_n) \right],$$

  where $B_n(\hat{\theta}_n) = B_n(y_1,...,y_n,\ x_1,...,x_n,\ \hat{\theta}_n)$ is calculated on the original sample, and $C_n^*(\hat{\theta}_n) = C_n(y_1^*,...,y_n^*, x_1^*,...,x_n^*, \hat{\theta}_n)$ is calculated on the bootstrap samples.
- **Approach B:** The second approach also re-calculates "the denominator":

$$\sqrt{n} B_n^*(\hat{\theta}_n) C_n^*(\hat{\theta}_n) \left[\frac{1}{n} \sum_{i=1}^{n} h_{in}(x_i^*, \hat{\theta}_n) \right],$$

  where $B_n^*(\hat{\theta}_n) = B_n(y_1^*,...,y_n^*, x_1^*,...,x_n^*, \hat{\theta}_n)$ and $C_n^*(\hat{\theta}_n) = C_n(y_1^*,...,y_n^*,$ $x_1^*,...,x_n^*, \hat{\theta}_n)$ are both calculated with the bootstrap samples.

Remark 1 Note that the Newton-Raphson bootstrap discussed in salibian-Barrera (2003) is a particular case of this approach, when $B_n(\theta)=[\nabla g_n(\theta)]^{-1}$, $C_n(\theta) = I_n$ and $g_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} h_{in}(y_i, x_i, \theta)$.

Remark 2 Note that in the case of robust linear regression approach "A" is the "robust bootstrap" introduced by Salibian-Barrera and Zamar (2002) . This inference method has two very desirable properties: its computational cost is low (and hence it is feasible in high dimensional problems) and it is robust to the presence of outliers in the data. These properties are particularly important when doing inference based on robust estimates.

4 Consistency of the Fast Bootstrap

In this section we show that the bootstrap distribution of the fast bootstrap converges weakly to the same limiting distribution as the estimates of interest. . . .

What I want to do

I want to use the Fast Bootstrap (Salibian-Barrera and Zamar, 2003) for the Cox regression model to estimate the sampling distribution of the partial likelihood estimator $\hat{\beta}$.

The proportional-hazards model (Cox, 1972) specifies that the hazard at time y for the ith individual whose covariate vector is x_i is given by

$$\lambda(y, x_i) = \lambda_0(y) e^{\beta^T x_i}$$

The parameter vector β is estimated by maximizing a partial likelihood. The so-called partial likelihood is

$$L_p(\beta) = \prod_u \frac{e^{\beta^T x_i}}{\sum_{j \in \Re_i} e^{\beta^T x_j}} \quad ,$$

where $\Re_i = \Re(y_i)$, be the risk set immediately prior to the ith failure, that is the set of individuals any of whom may be found to fail at time y_i .

Cox suggested that we treat his partial likelihood as an ordinary likelihood. In particular, to find the maximum likelihood estimate, use the score vector and the sample information matrix:

$$\frac{\partial}{\partial \beta} \log L_p(\beta) = \left(\frac{\partial}{\partial \beta_1} \log L_p(\beta), \ldots, \frac{\partial}{\partial \beta_p} \log L_p(\beta) \right)^T ,$$

$$i(\beta) = -\frac{\partial^2}{\partial \beta^2} \log L_p(\beta) ,$$

$$= -\begin{pmatrix} \frac{\partial^2}{\partial\beta_1\partial\beta_1}\log L_p(\beta) & \cdots & \frac{\partial^2}{\partial\beta_1\partial\beta_p}\log L_p(\beta) \\ \vdots & & \vdots \\ \frac{\partial^2}{\partial\beta_p\partial\beta_1}\log L_p(\beta) & \cdots & \frac{\partial^2}{\partial\beta_p\partial\beta_p}\log L_p(\beta) \end{pmatrix}.$$

We want to solve the equations

$$\frac{\partial}{\partial\beta}\log L_p(\beta) = 0,$$

which usually requires iterative methods. Therefore, if $\hat{\beta}^0$ is an initial guess, let

$$\hat{\beta}^1 = \hat{\beta}^0 + i^{-1}(\hat{\beta}^0)\frac{\partial}{\partial\beta}\log L_p(\hat{\beta}^0). \qquad (12)$$

where $i(\beta)$ is the sample information matrix. If $\hat{\beta}$ is the solution, Cox asserted

$$\hat{\beta} \overset{a}{\sim} N(\beta,\ i^{-1}(\beta)) .$$

<u>What I suggest according to Salibian-Barrera and Zamar (2003)</u>

From (12) a Taylor expansion around β gives

$$\hat{\beta} - \beta = A^{-1}(\beta)\frac{\partial}{\partial\beta}\log L_p(\beta) \qquad (13)$$

where

$$A(\beta) = \frac{\partial^2}{\partial\beta^2}\log L_p(\beta).$$

Equation (13) suggests the choices

$$h_{in}(y_i, x_i, \beta) = h(y_i, x_i, \beta) = \frac{\partial}{\partial\beta}\log L_p(\beta),$$

and

$$A(\beta) = \frac{\partial^2}{\partial \beta^2} \log L_p(\beta) .$$

Moreover, I want to compare the performance of the Fast Bootstrap method (Salibian-Barrera and Zamar, 2003) for the Cox regression model with that of the classical bootstrap, which requires of course computing power and time, using Melanoma data (Angelo Canty's example from his article on line "Resampling Methods in R: the boot Package, 2002").