MATH 523: CONDITIONAL LOGISTIC REGRESSION

The standard logistic regression model for binary responses sets

$$\pi_i(x_i) = \Pr[Y_i = y_i | x_i] = \frac{\exp\{y_i(\beta_0 + x_i\beta)\}}{1 + \exp\{(\beta_0 + x_i\beta)\}} \qquad y_i = 0, 1.$$

Computing the likelihood, $T_1(\boldsymbol{Y})$ (scalar) is sufficient for β_0, and $T_2(\boldsymbol{Y}, \mathbf{X})$ ($1 \times p$) is sufficient for β

$$L_n(\beta_0, \beta) = \frac{\exp\{T_1(\boldsymbol{y})\beta_0 + T_2(\boldsymbol{y}, \mathbf{x})\beta)\}}{\prod_{i=1}^{n} \{1 + \exp\{(\beta_0 + x_i\beta)\}\}} \qquad \text{where} \quad T_1(\boldsymbol{y}) = \sum_{i=1}^{n} y_i \qquad T_2(\boldsymbol{y}, \mathbf{x}) = \sum_{i=1}^{n} y_i x_i$$

Here, β_0 is a *nuisance* parameter; we are interested in the impact of the covariates on response, so β is the *parameter of interest*. Thus, consider the likelihood conditional on $T_1(\boldsymbol{Y}) = t_1$:

$$\Pr[\boldsymbol{Y} = \boldsymbol{y} | T_1 = t_1] = \frac{\Pr[Y_1 = y_1, \ldots, Y_n = y_n]}{\sum_{\mathbf{z} \in \mathcal{Y}(t_1)} \Pr[Y_1 = z_1, \ldots, Y_n = z_n]}$$

where $\mathcal{Y}(t) = \{(z_1, \ldots, z_n) : T_1(\mathbf{z}) = t\}$. We have from above that

$$\Pr[Y_1 = y_1, \ldots, Y_n = y_n] \equiv L_n(\beta_0, \beta) = \frac{\exp\{T_1(\boldsymbol{y})\beta_0 + T_2(\boldsymbol{y}, \mathbf{x})\beta)\}}{\prod_{i=1}^{n} \{1 + \exp\{(\beta_0 + x_i\beta)\}\}} = \frac{\exp\{t\beta_0 + T_2(\boldsymbol{y}, \mathbf{x})\beta)\}}{\prod_{i=1}^{n} \{1 + \exp\{(\beta_0 + x_i\beta)\}\}}$$

where, note, the denominator does not depend on $\boldsymbol{y}$. Therefore

$$\Pr[\boldsymbol{Y} = \boldsymbol{y} | T_1 = t_1] = \frac{\exp\{t\beta_0 + T_2(\boldsymbol{y}, \mathbf{x})\beta\} / \left\{\prod_{i=1}^{n} \{1 + \exp\{(\beta_0 + x_i\beta)\}\}\right\}}{\sum_{\mathbf{z} \in \mathcal{Y}(t_1)} \exp\{t\beta_0 + T_2(\mathbf{z}, \mathbf{x})\beta\} / \left\{\prod_{i=1}^{n} \{1 + \exp\{(\beta_0 + x_i\beta)\}\}\right\}} = \frac{\exp\{T_2(\boldsymbol{y}, \mathbf{x})\beta\}}{\sum_{\mathbf{z} \in \mathcal{Y}(t_1)} \exp\{T_2(\mathbf{z}, \mathbf{x})\beta\}}$$

This conditional distribution does not involve β_0. Therefore, we can consider the *conditional likelihood*

$$L_n^{(T_1)}(\beta; t_1) = \frac{\exp\{T_2(\boldsymbol{y}, \mathbf{x})\beta\}}{\sum_{\mathbf{z} \in \mathcal{Y}(t_1)} \exp\{T_2(\mathbf{z}, \mathbf{x})\beta\}} \tag{1}$$

and estimate β using it. This is a regular likelihood, so ML estimators have standard properties.

This argument can be taken further; suppose $\beta = (\beta_1, \ldots, \beta_p)^\mathsf{T}$, but that only β_1 is of interest. We can form the conditional likelihood conditioning on $T_1 = t_1$ **and** $T_{2j} = t_{2j}$ for $j = 2, \ldots, p$, where T_{2j} is the jth element of $T_2(\boldsymbol{y}, \mathbf{x})$. This leads to the following conditional likelihood for β_1:

$$L_n^{(T_1, T_{22}, \ldots, T_{2p})}(\beta_1; t_1, t_{22}, \ldots, t_{2p}) = \frac{\exp\{T_{21}(\boldsymbol{y}, \mathbf{x})\beta_1\}}{\sum_{\mathbf{z} \in \mathcal{Y}(t_1, t_{22}, \ldots, t_{2p})} \exp\{T_{21}(\boldsymbol{y}, \mathbf{z})\beta_1\}} \tag{2}$$

where $\mathcal{Y}(t_1, t_{22}, \ldots, t_{2p}) = \{(z_1, \ldots, z_n) : T_1(\mathbf{z}) = t_1, T_{2j}(\mathbf{z}, \mathbf{x}) = t_{2j}, j = 2, \ldots, p\}$. Again, conditional ML estimation of β_1 is relatively straightforward. Notice that testing the hypotheses $H_0 : \beta = 0_p$ using (1), or $H_0 : \beta_1 = 0$ using (2), all that is required is the probability distribution of the test statistics

$$L_n^{(T_1)}(0_p; t_1) \qquad \text{and} \qquad L_n^{(T_1, T_{22}, \ldots, T_{2p})}(0; t_1, t_{22}, \ldots, t_{2p}).$$

These are discrete distributions, so to compute p-values, we must compute the proportion of possible configurations of $\boldsymbol{y} = (y_1, \ldots, y_n)$ yielding test statistics at least as large as the observed values.

Conditional Logistic Regression for Matched Pairs or Stratified Data.
A *Matched Pair* study involves the measurement of pairs of responses; the pairing can be the result of making two observations on the same experimental unit (at time 1 and time 2, or before and after treatment etc.), or on matched (by age, sex or other characteristics) pairs of different experimental units. In the binary response case, let (Y_{i1}, Y_{i2}) be the responses for the ith matched pair, and suppose that $\pi_{ij} = \Pr[Y_{ij} = 1]$. Let $x_j = j - 1$, for $j = 1, 2$, and consider the model

$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \beta_{0i} + \zeta x_j = \begin{cases} \beta_{0i} & j = 1 \\ \beta_{0i} + \zeta & j = 2 \end{cases}$$

The parameters of the resulting model are the *pair-* or *subject-specific* intercepts, $\beta_{01}, \ldots, \beta_{0n}$, and the *subject-specific* log odds ratio, ζ.

Note that the interpretation of ζ is different from the interpretation of the *population* log odds ratio β_1 ($= \log\psi = \xi$ in previous notation) in the analysis of 2×2 tables for independent data using logistic regression; β is measures association at the population level, and arises as a population average across independent individuals, whereas ζ measures association at the individual matched pair level.

We can compute the full likelihood for these data as

$$L_n(\beta_{01}, \ldots, \beta_{0n}, \zeta) = \prod_{i=1}^{n} \left\{ \left(\frac{e^{\beta_{0i}}}{1+e^{\beta_{0i}}}\right)^{y_{i1}} \left(\frac{1}{1+e^{\beta_{0i}}}\right)^{1-y_{i1}} \left(\frac{e^{\beta_{0i}+\zeta}}{1+e^{\beta_{0i}+\zeta}}\right)^{y_{i2}} \left(\frac{1}{1+e^{\beta_{0i}+\zeta}}\right)^{1-y_{i2}} \right\}.$$

Then following the argument above, we can form the conditional likelihood to eliminate the nuisance parameters $\beta_{01}, \ldots, \beta_{0n}$, that is, we condition on $s_i = y_{i1} + y_{i2}$ for $i = 1, \ldots, n$. The random variables $S_1, \ldots, S_n$ each can take the values $0, 1, 2$. It can be easily shown that

$$\Pr[Y_{i1} = y_{i1}, Y_{i2} = y_{i2} | S_i = s] = \begin{cases} 1 & (s, y_{i1}, y_{i2}) = (0,0,0) \text{ or } (2,1,1) \\ \dfrac{e^{\zeta}}{1+e^{\zeta}} & (s, y_{i1}, y_{i2}) = (1,0,1) \text{ or } (1,1,0) \end{cases}$$

and therefore the conditional likelihood for ζ, given $\boldsymbol{S} = (s_1, \ldots, s_n)$ is

$$L_n^{(\boldsymbol{S})}(\zeta; s_1, \ldots, s_n) = \frac{e^{n_{01}\zeta}}{(1+e^{\zeta})^{n_{01}+n_{10}}}$$

where n_{01} is the number of observations with $y_{i1} = 0, y_{i2} = 1$, and n_{10} is the number of observations with $y_{i1} = 1, y_{i2} = 0$ (that is, the *discordant* pairs). It is easy to derive the ML estimate, and the Delta method gives the estimated standard error:

$$\widehat{\zeta} = \log\left(\frac{n_{01}}{n_{10}}\right) \qquad \text{e.s.e}(\widehat{\zeta}) = \sqrt{\frac{1}{n_{01}} + \frac{1}{n_{10}}}$$

whereas the population estimate is $\widehat{\xi} = \log(n_{00}n_{11}/n_{01}n_{10})$

This approach can be extended to handle *stratified* data, where each of K strata has its own intercept parameter but stratum k contains n_k individuals rather than a single matched pair,

$$\log\left(\frac{\pi_{ijk}}{1-\pi_{ijk}}\right) = \beta_{0k} + \zeta x_j \qquad i = 1, \ldots, n_k, j = 1, 2, k = 1, \ldots, K$$

and problems with multiple covariates.