Applied Statistical Modeling & Inference: Homework #5

Due on April 30, 2015 at 5:00pm

Professor Ying Lu & Professor Daphna Harel

Sriniketh Vijayaraghavan

Problem 1

Show that the Poisson distribution belongs to the Exponential family. For $Y \sim Poisson(\lambda)$, identify θ, $b(\theta)$ and $c(Y_i, \phi)$. Derive the mean and variance of Y in terms of θ, and further specify the variance as a function of the mean parameter and ϕ.

Solution
If $f(y_i;\theta_i;\phi)$ can be written in the form $exp[\frac{y.\theta - b(\theta)}{a(\phi)} + c(y,\phi)]$, then it is in the exponential family.
We know that for a Poisson distribution of the form $Y \sim Poisson(\lambda)$, the PMF is

$$\frac{\lambda^x}{x!}.e^{-\lambda}$$

This can be converted as follows:

$$exp\{x.log(\lambda) - \lambda - log(x!)\}$$

So, we see that $\boxed{y = x, \theta = log(\lambda), b(\theta) = e^\theta, \phi = 1, c(y,\phi) = -log(x!)}$.
We also know that

$$I = log(f(y,\theta,\phi)) = \frac{(y.\theta - b(\theta))}{\phi} + c(y,\phi)$$
$$\frac{\partial I}{\partial \theta} = \frac{(y - b'(\theta))}{\phi}$$
$$\frac{\partial^2 I}{\partial \theta^2} = \frac{b''(\theta)}{\phi}$$

From the first Bartlett Identity, we know that

$$E(\frac{\partial I}{\partial \theta}) = 0 = \frac{(E(y) - b'(\theta))}{\phi}$$

$$\boxed{\implies E(y) = b'(\theta) = e^\theta}$$

From the Second Bartlett Identity we know that

$$E(\frac{\partial^2 I}{\partial \theta^2}) + E((\frac{\partial I}{\partial \theta})^2) = 0$$
$$\implies \frac{-b''(\theta)}{\phi} + E(\frac{(y - b'(\theta))^2}{\phi^2}) = 0$$
$$\frac{-b''(\theta)}{\phi} + \frac{V(Y)}{\phi^2} = 0$$
$$\implies V(Y) = b''(\theta).\phi$$

Thus,

$$\boxed{V(Y) = e^\theta}$$

Let $E(Y) = \mu = b'(\theta)$, then we can express $V(Y)$ in terms of μ

$$\boxed{V(Y) = b''(\theta).\phi = b''(b'^{-1}(\mu)).\phi = V(\mu).\phi}$$

Problem 2

Derive the Maximum Likelihood Estimator of a Poisson Regression: $Y_i \sim Poisson(\lambda_i)$, $log(\lambda_i) = X_i^T\beta, i = 1, ..., n$, assuming common offset value for each observation.

- Find sufficient statistic for β.
- Show that $log(\lambda)$ is a canonical link.
- Derive the score vector and the Hessian Matrix, and express them in terms of X, y and the Fisher Scoring Algorithm to obtain $\hat{\beta}_{MLE}$.

Solution

a. We know that the PMF of a poisson distribution is:

$$f(y|\lambda) = \frac{e^{-\lambda}.\lambda^y}{y!}, y = 0, 1, 2, ...$$

$$L(\lambda|y) = \prod_{i=1}^{N} \frac{e^{-\lambda}.\lambda^{y_i}}{y_i!}$$

The likelihood function is

$$l(\lambda|y) = \sum_{i=1}^{N} log(\frac{e^{-\lambda}.\lambda^{y_i}}{y_i!}) = \sum_{i=1}^{N} -\lambda + y_i log(\lambda) - log(y_i)$$

$$\frac{\partial l(\lambda|y)}{\partial \lambda} = \sum_{i=1}^{N} -1 + \frac{y_i}{\lambda} = 0$$

$$-N + \frac{1}{\lambda}\sum_{i=1}^{N} y_i = 0$$

$$\hat{\lambda} = \frac{1}{N}\sum_{i=1}^{N} y_i$$

Thus, the MLE of $\hat{\lambda}$ is the average of $y_1, ..., y_N$. We know since the expectation of Poisson variable is $E[y] = \lambda$.

b. If a link function is chosen such that $\theta = X_i^T\beta$, then

$$\sum_{i=1}^{N} Y_i\theta = \sum_{i=1}^{N} Y_i X_i^T\beta = \sum_{i=1}^{N} Y_i \sum_{j=1}^{p} X_{ij}\beta_j = (X^TY)\beta,$$

then X^TY is a sufficient statistic for β.

c. If the link function is chosen such that $\theta = X_i^T\beta$, then it is called a canonical link function. Since we know that it is a poisson regression, we know that

$$log(\mu) = X\beta$$

In logistic regression, we identified logit as a 'canonical link' because

$$g'(\lambda) = \frac{1}{\lambda}$$

$$\implies g(\lambda) = log(\lambda)$$

Problem 2 continued on next page. . .

d.

$$I = \sum_{i=1}^{N} \frac{(y_i\theta_i - b(\theta_i))}{\phi} + \sum_{i=1}^{N} c(y_i, \phi), where$$

$\mu_i = b'(\theta_i)$

$$g(\mu_i) = \eta_i = X_i^T \beta$$

score statistic $u_j = \frac{\partial I}{\partial \beta} = \sum_{i=1}^{n} \frac{\partial I}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_i}$, for $j = 1, ..., p$. Applying the chain rule for $\eta = X\beta, \mu = g^{-1}(\eta), \theta = b'^{-1}(\mu)$ we have the estimating equation

$$u_j = \frac{y-\mu}{\phi} V(\mu) \frac{\partial \mu}{\partial \eta} x_j = (y-\mu) W \frac{\partial \eta}{\partial \mu} x_j = 0$$

where W is a nxn diagonal weight matrix,

$$W = [diag(g'(\mu)^2 V(\mu_1)\phi, ..., g'(\mu_n)^2 V(\mu_n)\phi)]^- 1.$$

Or equivalently,

$$\sum_{i=1}^{n} w_{ii}(y_i - \mu_i) \frac{\partial \eta_i}{\partial \mu_i} x_{ij} = 0$$

The Hessian H is a pxp matrix, with elements (h_{jk})

$$h_{jk} = \sum_{i=1}^{n} \frac{\partial u_j}{\partial \beta_k}$$

Writing it in matrix form,

$$H = X^T (y-\mu) [\frac{\partial W}{\partial \beta_k} \frac{\partial \eta}{\partial \mu}] - X^T W X$$

e. From the Newton-Raphson method, we know that,

$$\hat{\beta}_{(I+1)} = \hat{\beta}_{(I)} - H^{-1}(\hat{\beta}_{(I)} U(\hat{\beta}_{(I)}))$$

The Fisher Scoring algorithm replaces the observed information $-H$ by the information $I = -E(H)$. This quantity is positive definite

$$\hat{\beta}_{(I+1)} = \hat{\beta}_{(I)} + I^{-1}(\hat{\beta}_{(I)} U(\hat{\beta}_{(I)}))$$

Problem 3

In this part, you will use a dataset from Long(1990) on the number of publications produced by PhD biochemists. There are following variables in the dataset:

- art: articles in the last three years of PhD
- fem: female = 1, male = 0
- mar: coded on if married
- kid5: number of children under age six
- phd: prestige score of PhD program
- ment: articles by mentor in three years

The focus of this question is on how to fit a Poisson model and interpret the coefficients (rather than select the best model):

- First fit a Poisson Regression model that include all the available covariates (additive effects only) in R using glm command, with log link function. Report the results of the model, and interpret the effects of the covariates in terms of Risk Ratio. Report 95% confidence interval for the Risk Ratio and comment on their statistical significance at level 5%.
- Fir the model using the Fisher Scoring algorithm that you developed in the previous question in R.
  - You can generate the design matrix X using model.matrix function.
  - The initial values can be found running OLS of Y on X.

  Estimate the observed information matrix, and explain briefly that under Poisson Regression, that is also the Information Matrix.
- Estimate the standard errors for β_{MLE}.
- Reproduce other statistics (such as Deviance) from the glm output.

a.

Problem 4

In this part, you are asked to improve the Logistic regression model on children's attendance in religious school that we studied in class. Use dataset religion1.csv

- The covariates are: sex (1 = female, 0 = male); edu (1 = college or above; 0 = below college); age, agesq(age squared); income level (in $10,000 brackets); attend(frequency of attending religious services); race(1 = white, 0 = nonwhite); Married(Yes and No)

Pay particular attention to the following issues:

- The candidate models include all additive and two-way interactions between any two pairs of variables.
  - Use the formula command and write a loop to generate all the possible additive models. (See Lab 10 example)
  - For variables income in $10,000 brackets, think carefully how you can incorporate those variables respecting the parsimony principle.
  - Based on the additive model(s) you select, build in two-way interactions.
- For each model, the log-likelihood, deviance, df, AIC, BIC, pseudo-R^2, classification rate, sensitivity and specificity.
- Choose the best model and report the results, make sure you interpret the effects of the covariates in terms of odds ratio. (The benchmark model has AIC = 369.16, BIC = 395.41, to get full credit, your best model should be comparable to this performace)
- For the model of your choice, conduct an ROC analysis and plot the ROC curve, estimate the area under the curve using a stepwise function approximation.
- **Bonus** Best competition model: Try to build a model that outperforms the benchmark model. The student who builds the best model gets an extra point. For any other models that outperforms the benchmark model, you will get a bonus point that is proportional to the improvement you accomplish relative to the best model.

a.