December 13, 2018

STAT 545: Solutions to Homework 4

Dennis D. Cox Department of Statistics Rice University

December 13, 2018

1 Exercise 4.5

(a) Note that the R-function glm automatically gives the null deviance (fitting only the intercept), so it is not necessary to fit this separately. According to the help on the glm function, the deviance is

up to a constant, minus twice the maximized log-likelihood. Where sensible, the constant is chosen so that a saturated model has deviance zero.

In this case, with this grouped data, the value of x = 2 has all successes, so the saturated model deviance is not defined. So, if I understand the help file, the deviances that are returned are just minus 2 times the maximized log likelihood, so L_0 is -1/2 times the null.deviance returned by R, and L_1 is -1/2 times the deviance. That seems easy enough to compute. Here is the R session:

```
> xgrouped = c(0,1,2)
> ygrouped = matrix(c(1,2,4,3,2,0),ncol=2)
> ygrouped
      [,1] [,2]
[1,] 1 3
[2,] 2 2
[3,] 4 0
> M1grouped = glm(ygrouped ~ xgrouped,family=binomial)
```

```
> M1grouped$deviance
[1] 0.9843993
> M1grouped$null.deviance
[1] 6.25678
> M1grouped$null.deviance - M1grouped$deviance
[1] 5.27238
```

OK, let's check by doing a direct calculation of the log maximized likelihoods:

```
> L1grouped = sum(dbinom(ygrouped[,1],c(4,4,4),
+ plogis(M1grouped$linear.predictors),log=TRUE))
> L1grouped
[1] -2.336075
```

(Note that *plogis* is the CDF of the standard logistic distribution, is the inverse of the logit link function.) Whoops. That doesn't agree with the value of -0.4921997 in the previous calculation. I am sure (other than I may have made a typing error) that the -2.336075 value is correct, but why the difference? Maybe they included the "saturated model" terms for the $-2 \log$ likelihood on a term by term basis. Thus, it makes sense for the values of x except x = 2. This is my best guess. Let's see if it is correct:

```
> M1grouped.deviance = 2*(-L1grouped + sum(dbinom(ygrouped[1:2,2],
+ c(4,4),ygrouped[1:2,2]/4,log=TRUE)))
> M1grouped.deviance
[1] 0.9843993
```

OK, so that agrees with the deviance we found before. We should be able to do the same with the null deviance. Under the null model of a constant probability of success, independent of x, the m.l.e. of π is just the total proportion of successes.

```
> L0grouped = sum(dbinom(ygrouped[,1],c(4,4,4),
+ sum(ygrouped[,1])/12,log=TRUE))
> L0grouped
[1] -4.972265
> M0grouped.deviance = 2*(-L0grouped + sum(dbinom(ygrouped[1:2,2],
+ c(4,4),ygrouped[1:2,2]/4,log=TRUE)))
> M0grouped.deviance
[1] 6.25678
```

This agrees with the value returned by the glm function. Now we turn to the ungrouped data.

```
> # Now doing the ungrouped data
> xungrouped = rep(c(0,1,2),c(4,4,4))
> yungrouped = c(1,0,0,0,1,1,0,0,1,1,1,1)
> M1ungrouped = glm(yungrouped ~xungrouped,family=binomial)
> M1ungrouped$deviance
[1] 11.02826
> M1ungrouped$null.deviance
[1] 16.30064
> M1ungrouped$null.deviance - M1ungrouped$deviance
[1] 5.27238
```

For this ugrouped data, the "saturated" model is not sensible for all values of x since the (single) y value is either 0 or 1. Therefore, we conjecture that the deviance is just the -2 times the maximized log likelihood. Checking this out:

```
> L1ungrouped = sum(dbinom(yungrouped,rep(1,12),
+ plogis(M1ungrouped$linear.predictors),log=TRUE))
> -2*L1ungrouped
[1] 11.02826
```

So that's correct.

So, in a table, here are our results:

	grouped	ungrouped
LO	-3.12839	-8.150319
L1	-0.4921997	-5.514129

According to the claim in the exercise, we should see the same numbers in each row of the table. Note that the difference is the same within each row (5.021929). This is because the binomial coefficients were included in computing the grouped log likelihood, whereas they are all 1 in the ungrouped case. I guess the textbook author forgot about them.

(b) The deviances are different because there are different disallowed terms in the saturated model. For the grouped data, we just leave off the term corresponding to x = 2, while for the ungrouped data all terms in the saturated model don't make sense are hence are left out when computing the

deviances. Note that the binomial coefficients don't matter for the deviances since they are the same for the given and saturated model so they drop out, basically.

(c) Yes, the difference in deviances is 5.27238 for both ways of entering the data.

The computed p-value for the LRT statistic for testing H_0 : $\beta = 0$ is 0.02166637, but the accuracy of the χ_1^2 approximation is very dubious as there are only N = 12 observations. Note that the p-value from the Wald test is 0.0683, and the Wald and LRT are asymptotically equivalent in this case, so clearly the asymptotic results are not accurate.

2 Exercise 4.18

When all the success probabilities in the binomials are the same, then we may add them together to get a single $Binom(n,\pi)$ observation with $n = \sum_i n_i$. The MLE for π is just the total number of successes over the total number of trials as given as mentioned on p. 13 at the beginning of section 1.4.

The $N \times 2$ table is constructed like the grouped data before: put (say) y_i in the first column and $n_i - y_i$ in the second column. To test the null hypothesis, we have

$$\hat{\mu}_{i1} = n_i \hat{\pi} \hat{\mu}_{i2} = n_i (1 - \hat{\pi})$$

Thus, Pearson's chi-squared is given by

$$\Xi^{2} = \sum_{i} \left[\frac{(y_{i} - n_{i}\hat{\pi})^{2}}{n_{i}\hat{\pi}} + \frac{(n_{i} - y_{i} - n_{i}(1 - \hat{\pi}))^{2}}{n_{i}(1 - \hat{\pi})} \right]$$
$$= \sum_{i} \frac{(y_{i} - n_{i}\hat{\pi})^{2}}{n_{i}\hat{\pi}(1 - \hat{\pi})}.$$

If all $n_i = 1$ then this is just

$$\frac{1}{\hat{\pi}(1-\hat{\pi})} \sum_{i} (y_i - \hat{\pi})^2 = \frac{1}{\hat{\pi}(1-\hat{\pi})} \left[\sum_{i} y_i - 2\hat{\pi} \sum_{i} y_i + N\hat{\pi}^2 \right]$$
$$= \frac{1}{\hat{\pi}(1-\hat{\pi})} \left[N\hat{\pi} - N\hat{\pi}^2 \right]$$
$$= N.$$

Note that $y_i^2 = y_i$ in this case since y_i is either 0 or 1.

3 Exercise 5.8

I'll just do parts (a) and (b) in one pass. Here is the R-code:

```
> x1 = c(12,15,42,52,59,73,82,91,96,105,114,120,121,128,130,139,139,157)
> x0 = c(1,1,2,8,11,18,22,31,37,61,72,81,97,112,118,127,131,140,151,159,177,206)
> x = c(x0, x1)
> length(x0)
[1] 22
> length(x1)
[1] 18
> y = c(rep(0,22),rep(1,18))
> plot(x,y,main="Plot for Exercise 5.8")
> fit1 = glm(y ~ x ,family=binomial)
> summary(fit1)
Call:
glm(formula = y ~ x, family = binomial)
Deviance Residuals:
    Min
                   Median
                                ЗQ
              1Q
                                        Max
-1.3126 -1.0907 -0.9482
                            1.2170
                                      1.4052
Coefficients:
             Estimate Std. Error z value Pr(|z|)
(Intercept) -0.572693
                        0.602395
                                 -0.951
                                             0.342
             0.004296
                        0.005849
                                   0.734
                                             0.463
х
(Dispersion parameter for binomial family taken to be 1)
    Null deviance: 55.051
                           on 39 degrees of freedom
Residual deviance: 54.504 on 38 degrees of freedom
AIC: 58.504
Number of Fisher Scoring iterations: 4
```

```
> # slope is not significant
> fit2 = glm(y \sim x + x\sim2, family = binomial)
> summary(fit2)
Call:
glm(formula = y ~ x + x^2, family = binomial)
Deviance Residuals:
                               ЗQ
    Min
             1Q Median
                                        Max
-1.3126 -1.0907 -0.9482 1.2170 1.4052
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.572693
                       0.602395 -0.951
                                           0.342
                       0.005849 0.734
             0.004296
                                            0.463
х
(Dispersion parameter for binomial family taken to be 1)
    Null deviance: 55.051 on 39 degrees of freedom
Residual deviance: 54.504 on 38 degrees of freedom
AIC: 58.504
Number of Fisher Scoring iterations: 4
> x2 = x^{2}
> fit2 = glm(y ~ x + x2, family = binomial)
> summary(fit2)
Call:
glm(formula = y ~ x + x2, family = binomial)
Deviance Residuals:
   Min
            1Q Median
                            ЗQ
                                  Max
-1.482 -1.009 -0.507 1.012
                                 1.788
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
```

```
7
```

(Intercept) -2.0462547 0.9943478 -2.0580.0396 * 2.242 0.0600398 0.0267808 0.0250 * x x2 -0.0003279 -2.097 0.0001564 0.0360 * 0 *** 0.001 ** 0.01 * 0.05 . 0.1 Signif. codes: 1 (Dispersion parameter for binomial family taken to be 1) Null deviance: 55.051 on 39 degrees of freedom Residual deviance: 48.228 on 37 degrees of freedom AIC: 54.228 Number of Fisher Scoring iterations: 4 > # both coefficients of x and x² are significant starting httpd help server ... done > lines(0:206,plogis(-2.0462547 + 0.0600398*(0:206) - 0.0003279*(0:206)^2))

The plot appears somewhere below.

Note that when we just fit the linear model, the coefficient is not significant. Clearly this is because there are two many zeroes (for y values) on either end and not enough in the middle. The quadratic model fixes this nicely, and both coefficients are significant! We see that the risk of kyphosis is low for both the very young and the older patients and it is the patients of intermediate age (between about 4 and 12) who are at highest risk.

(Question: what was wrong with the original fit2 statement?)

4 Exercise 5.11

This question is rather confusing. So, I'm assuming the outcome variable y is 1 if the person uses oral contraception and 0 if the person doesn't. Looking at the "Education" variable, there is a positive coefficient meaning that higher education (≥ 1 year of college) implies higher odds of using oral contraception. I think when it says "conditional odds ratio" it means the ratio of the odds of using oral contraception conditional on education at the the higher level of education over the odds at the lower level of education, conditional on everything else (Age, Race, and Marital Status). Let the

predictor variable indicators be labeled A, R, M, and E (first letters) with values 1 is indicated in the table. Let α denote the intercept. Then the fitted model is

$$\log \frac{P[Y=1|A, R, M, E]}{P[Y=0|A, R, M, E]} = \alpha - 1.320A + 0.622R - 0.460M + 0.501E.$$

I think the (estimated) conditional odds ratio asked for is

$$\begin{aligned} &P[Y = 1|A, R, M, E = 1] / P[Y = 0|A, R, M, E = 1] \\ &P[Y = 1|A, R, M, E = 0] / P[Y = 0|A, R, M, E = 0] \\ &= \exp\left[(\alpha - 1.320A + 0.622R - 0.460M + 0.501) - (\alpha - 1.320A + 0.622R - 0.460M)\right] \\ &= \exp\left[0.501\right]. \end{aligned}$$

Thus, we can get the approximate confidence interval for the coefficient from the data in the table and exponentiate to get the confidence interval for the desired odds ratio. It doesn't say what level of confidence to use, so I will choos 95%.

$$0.501 \pm 1.96 * 0.077 = (0.652, 0.350).$$

 $e^{.652} = 1.92$
 $e^{.350} = 1.42.$

So our confidence interval is $1.42 \leq \text{Odds Ratio} \leq 1.92$ with (approximately) 95% confidence.

5 Exercise 5.30

(a) The derivation is

$$\begin{split} \log \frac{P[Y=1|X=x]}{P[Y=0|X=x]} \\ &= \log \frac{f(x|Y=1)P[Y=1]/\{f(x|Y=1)P[Y=1] + f(x|Y=0)P[Y=0]\}}{f(x|Y=0)P[Y=0]/\{f(x|Y=1)P[Y=1] + f(x|Y=0)P[Y=0]\}} \\ &= \log \frac{f(x|Y=1)P[Y=1]}{f(x|Y=0)P[Y=0]} \\ &= \log \frac{P[Y=1]}{P[Y=0]} + \log \frac{(2\pi\sigma^2)^{-1/2}\exp[-(x-\mu_1)^2/(2\sigma^2)]}{(2\pi\sigma^2)^{-1/2}\exp[-(x-\mu_0)^2/(2\sigma^2)]} \\ &= \log \frac{P[Y=1]}{P[Y=0]} + x \left(\frac{\mu_1 - \mu_0}{\sigma^2}\right) - \frac{\mu_1^2 - \mu_0^2}{2\sigma^2} \\ &= \alpha + \beta x, \\ \alpha &= \log \frac{P[Y=1]}{P[Y=0]} - \frac{\mu_1^2 - \mu_0^2}{2\sigma^2} \\ \beta &= \frac{\mu_1 - \mu_0}{\sigma^2}. \end{split}$$

The first equation follows from Bayes formula (used in both the numerator and denominator). The second is a formula already discussed in class (posterior odds equals prior odds times the likelihood ratio, which is f(x|Y = 1)/f(x|Y = 0)). In the third equation we plugged in the normal densities. The rest is just algebra.

(b) We can use the calculation above and just plug in the different normal densities at the third equation.

$$\log \frac{P[Y=1|X=x]}{P[Y=0|X=x]}$$

$$= \log \frac{P[Y=1]}{P[Y=0]} + \log \frac{(2\pi\sigma_1^2)^{-1/2}\exp[-(x-\mu_1)^2/(2\sigma_1^2)]}{(2\pi\sigma_0^2)^{-1/2}\exp[-(x-\mu_0)^2/(2\sigma_0^2)]}$$

$$= \log \frac{P[Y=1]}{P[Y=0]} - \log \frac{\sigma_1}{\sigma_0} - x^2 \left(\frac{1}{2\sigma_1^2} - \frac{1}{2\sigma_0^2}\right) + x \left(\frac{\mu_1}{\sigma_1^2} - \frac{\mu_0}{\sigma_0^2}\right) - \left(\frac{\mu_1^2}{s\sigma_1^2} - \frac{\mu_0^2}{s\sigma_0^2}\right)$$

$$= \alpha + \beta_1 x + \beta_2 x^2,$$

The reader can easily see the expressions for the coefficients in this quadratic functions.

(c) Similarly to (b), we now plug in the densities from the exponential family as given

$$\log \frac{P[Y=1|X=x]}{P[Y=0|X=x]}$$

$$= \log \frac{P[Y=1]}{P[Y=0]} + \log \frac{\alpha(\theta_1)b(x)\exp[xQ(\theta_1)]}{\alpha(\theta_0)b(x)\exp[xQ(\theta_0)]}$$

$$= \alpha + \beta x,$$

$$\alpha = \log \frac{P[Y=1]}{P[Y=0]} + \log \frac{\alpha(\theta_1)}{\alpha(\theta_0)}$$

$$\beta = Q(\theta_1) - Q(\theta_0).$$

(d) This one is a slightly more complicated because we need to deal with the multivariate normal density. Let p denote the dimension of X. $|\Sigma|$ denotes the determinant of the covariance matrix.

$$\log \frac{P[Y=1|X=x]}{P[Y=0|X=x]}$$

$$= \log \frac{P[Y=1]}{P[Y=0]} + \log \frac{(2\pi)^{-p/2} |\Sigma|^{-1/2} \exp[-(x-\mu_1)^T \Sigma^{-1}(x-\mu_1)]}{(2\pi)^{-p/2} |\Sigma|^{-1/2} \exp[-(x-\mu_0)^T \Sigma^{-1}(x-\mu_0)]}$$

$$= \alpha + \beta^T x,$$

$$\alpha = \frac{P[Y=1]}{P[Y=0]} - \frac{1}{2} (\mu_1 - \mu_0)^T \Sigma^{-1} (\mu_1 - \mu_0)$$

$$\beta = \Sigma^{-1} (\mu_1 - \mu_0).$$



Plot for Exercise 5.8

Figure 1: Plot of data and quadratic logit fit.