Abbreviated Notes on Logistic and Poisson Regression

7.1 Motivation for Logistic Regression

Purpose: to perform regression-type analyses (including ANOVA-type analyses) when the response is either binary or proportions.

Motivation: Behavioral responses are sometimes either correct or incorrect. We may wish to consider probability of correct response as a function of some explanatory variable(s) or across conditions (or both). This arises particularly in psychophysical contexts.

Example: We used logistic regression to analyze saccade errors in hemispatial neglect patients. (We verified that errors increase when a saccade is to the left of eye fixation.)

Example (Psychophysics): Human observers were shown weak flashes of light by Hecht et al. (1942). Percent detection increases with the intensity. Similar experiments were performed by Hartline et al. (1947) on the horseshoe crab Limulus; here the response was the proportion of trials on which an isolated optic nerve fiber fired.

Example (Psychophysical, I guess, though doesn’t sound like high Science): Cows were exposed to electrical current and responses (yes/no mouth movement) were recorded. The current values were 1,2,3,4,5 milliamps and the corresponding number of positive responses out of 70 were 9,21,47,60,63. (The proportions are .129,.300,.671,.857,.900.)

The big problem with linear regression is illustrated in the first figure: the proportions vary between 0 and 1 but a line $y = a + bx$ is unrestricted and thus can not represent the variation accurately (at least not for proportions that get close to 0 or 1). The problem is even more severe in the binary case, where the data are either 0 or 1.
Figure 1: Observed proportion of responses following application of electric probe together with fit from linear regression.

What is needed is a sigmoidal curve, which goes to zero as the explanatory variable $x$ goes to $-\infty$ and increases to one as the $x \to \infty$. One of the simplest such functions is $f(x) = \exp(x)/(1 + \exp(x))$. In logistic regression the line $y = a + bx$ is replaced by

$$y = \frac{\exp(a + bx)}{1 + \exp(a + bx)}.$$

The logistic regression solution is illustrated in the second figure.

## 7.2 The Logistic Regression Model

In linear regression we have $Y = \beta_0 + \beta_1 x_i + \epsilon_i$ and $\epsilon_i \sim N(0, \sigma^2)$. 

In logistic regression for Binomial data we have

$$Y_i \sim B(n_i, p_i)$$

$$p_i = \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)}.$$

Note problems in applying linear regression model to proportions: (i) line isn’t constrained to (0,1), (ii) variances are not equal, (iii) proportions are
not Normal (unless we have large samples). The logistic regression model fixes these.

To interpret the model we use the logit transformation:

$$\log \frac{p_i}{1-p_i} = \beta_0 + \beta_1 x_i.$$  

This says that the log odds (of a response) are linear in $x$. Thus, $\beta_1$ is the change in the log odds for a unit change in $x$.

The model extends immediately to multiple explanatory variables: we write

$$\log \frac{p_i}{1-p_i} = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_K x_{Ki}.$$  

### 7.3 Estimation and Testing

#### 7.3.1 Estimation

The coefficients are estimated by ML. The variance matrix and standard errors are obtained using the observed information matrix.
7.3.2 Hypothesis Tests

In general, we have the Likelihood Ratio Test:

\[ \Lambda = -2[\log(\hat{L}_1) - \log(\hat{L}_2)] \]

where \( \hat{L}_i \) is the maximum value of the likelihood under model \( i \).

Asymptotically \( \Lambda \) follows the \( \chi^2 \) distribution with degrees of freedom equal to the number of additional parameters in the larger model (\#2). (Nesting means that model \#1 is model \#2 with at least 1 specific parameter set to zero.)

Here, Nested Test of “Null Deviance” vs. “Deviance” is analogous to Overall F test

- Deviance is \(-2\log(\hat{L})\)
- Null Deviance is for “intercept-only” model
- The test is based on the asymptotic chi-squared distribution (of null.deviance-deviance). Degrees of freedom is equal to the number of covariates.

7.3.3 S Output

Compare the usual \texttt{lm} summary to the \texttt{glm} summary:

Coefficients:

|            | Value | Std. Error | t value | Pr(>|t|) |
|------------|-------|------------|---------|----------|
| (Intercept)| -0.0586 | 0.1020 | -0.5741 | 0.6061   |
| current    | 0.2100  | 0.0308 | 6.8268  | 0.0064   |

Residual standard error: 0.09728 on 3 degrees of freedom
Multiple R-Squared: 0.9395
F-statistic: 46.6 on 1 and 3 degrees of freedom, the p-value is 0.006431

... 

Coefficients:

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>Std. Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-2.973571</td>
<td>0.3520122</td>
<td>-8.447352</td>
</tr>
<tr>
<td>current</td>
<td>1.143587</td>
<td>0.1188744</td>
<td>9.620121</td>
</tr>
</tbody>
</table>

(Dispersion Parameter for Binomial family taken to be 1)
Null Deviance: 147.2314 on 4 degrees of freedom

Residual Deviance: 3.455919 on 3 degrees of freedom

Note that the difference in deviance is

\[ \text{null deviance} - \text{residual deviance} = 147.2 - 3.5 = 143.7 \]

which should be compared to the chi-squared distribution on 1 degree of freedom. It is very highly significant.

We interpret the fitted model as saying that, on average, for every 1 milliamp increase in current there is a 1.14 ± .13 increase in the log odds of a response. An approximate 95% CI for the factor by which the odds increase is \( \exp(1.14 ± 2(.13)) \), i.e., (2.4,4.1). When the probability is .5 this would imply an additional 1 milliamp would increase the probability to (.71, .80), with 95% confidence.

### 7.4 Implementation in S/R

```r
> expit_function(x) {exp(x)/(1+exp(x))}
> resp_c(9,21,47,60,63)
> noresp_70-resp
> prop_resps/70
> youch_cbind(resp,noresp)
> current_1:5
> cattle.df.data.frame(current,youch)
> cattle.lm.lm(prop~current)
> cattle glm glm(youch~current,family=binomial,data=cattle.df)
> x_seq(.1,5.9,by=.01)
> y1_cattle.lm$coef[1]+cattle.lm$coef[2]*x
> y2_expit(cattle glm$coef[1]+cattle glm$coef[2]*x)
> plot(current,prop,xlim=c(0,6),ylim=c(0,1))
> lines(x,y1)
> lines(x,y2)
> summary(cattle.lm)
> summary(cattle glm)
```

Note that when the data are binary a single column is sufficient: we do not need the \( n_i \) values (or, equivalently, the \( n_i - y_i \) values we use in S-PLUS) because they are all equal to 1. Thus, in using logistic regression for binary data in S-PLUS the response variable becomes a single column rather than a matrix of two columns.
7.5 Technical Comments

1. The \textit{expi} function \(\exp(x)/(1 + \exp(x))\) is one of many possible sigmoidal curves and thus logistic regression is only one of many possible models for binary or proportion data. In fact, the \textit{expi} is actually the cdf of the logistic distribution and the cdf of any continuous distribution would work instead. One important alternative is the Probit regression model, which substitutes the Normal cdf in place of the \textit{expi}. This plays a predominant role in signal detection theory.

2. Two things are special about the logistic regression model. First, it gives a nice interpretation of the coefficients in terms of log odds. Second, in the logistic regression model (but not the Probit or other versions) the loglikelihood function is concave. This means that there is a unique MLE, which can be obtained from an arbitrary starting value in the iterative algorithm. There are additional technical statistical special features of the logistic regression model. All in all, it is the usual method.

3. Residual analysis may be performed using “deviance residuals” (or other forms of residuals, including Pearson chi-squared residuals).

7.6 Motivation for Poisson Regression

Useful when we have counts depending on one or more explanatory variables.

Usually don’t have quite the same issue as with logistic regression, but here the count rate does have to be positive and the ordinary regression line is not constrained, and will eventually go negative. The simple solution is to use a log transformation of the underlying rate.

Example: SEF data Homework 4; ANOVA worked fine, but the right way to do ANOVA with Poisson data is to use Poisson regression.

Example: In smoothing the PSTH I like to apply Poisson regression.

7.7 The Poisson Regression Model

Here we have

\[
Y_i \sim P(\lambda_i) \\
\lambda_i = \exp(\beta_0 + \beta_1 x_i).
\]
To interpret the model we use the log transformation:

\[
\log \lambda_i = \beta_0 + \beta_1 x_i.
\]

### 7.8 Estimation and Testing

As in logistic regression we use ML estimation and the likelihood ratio test ("analysis of deviance").

Here I reanalyze the SEF data from Homework 4. I compare analysis of variance summary with linear regression summary, and then with Poisson regression summary.

```r
> sef.df
  sef cond
  1  9  L
  2  6  L
  3  9  L
  4  9  L
  5  6  L
  6  6  L
  7  8  L
  8  5  L
  9  7  L
 10  9  L
 11  4  L
 12  8  L
 13  8  L
 14  3  L
 15  6  L
 16  2  U
 17  0  U
 18  6  U
...
```

```r
> summary(aov(sef~cond, sef.df))
                   Df Sum of Sq  Mean Sq F value Pr(>F)
cond                3  228.8215  76.27384 18.75505 1.836788e-08
Residuals         54  219.6095   4.06684
> summary(lm(sef~cond, sef.df))
> 
    Call: lm(formula = sef ~ cond, data = sef.df)
    Residuals:
```

7
```
Min  1Q  Median    3Q   Max
-3.867  -1.643  -0.231  1.439  5.467

Coefficients:
              Value  Std. Error   t value  Pr(>|t|)
(Intercept)  3.4929      0.2650    13.1828   0.0000
  cond1      2.1119      0.3747     5.6362   0.0000
  cond2     -0.7405      0.2138    -3.4629   0.0011
  cond3     -0.5214      0.1547    -3.3701   0.0014

Residual standard error: 2.017 on 54 degrees of freedom  
Multiple R-Squared: 0.5103  
F-statistic: 18.76 on 3 and 54 degrees of freedom, the p-value is 1.837e-08

> summary(glm(sef~cond,sef.df,family=poisson))  
Call: glm(formula = sef ~ cond, family = poisson, data = sef.df)  
Deviance Residuals:
              Min      1Q  Median      3Q      Max
-2.299068 -1.158444 -0.1436831  0.7763143  2.732243

Coefficients:
              Value  Std. Error   t value
(Intercept)  1.1212194      0.07897981  14.196279
  cond1      0.4774091      0.09581966     4.982371
  cond2     -0.1732444      0.06277747    -2.759659
  cond3     -0.1548059      0.05231702    -2.958998

(Dispersion Parameter for Poisson family taken to be 1 )

Null Deviance: 149.7523 on 57 degrees of freedom

Residual Deviance: 92.48045 on 54 degrees of freedom

As in logistic regression we look at difference in deviance:
null.deviance - residual deviance = 149.8 - 92.5 =
which should be compared to the chi-squared distribution on 3 degrees of freedom. It is very highly significant.

7.9 Generalized Linear Models

Logistic and Poisson regression are special cases of generalized linear models. These generalize linear models by allowing a certain class of distribu-
```
tions (known as exponential families) that includes several common examples. They also use a link function that links the expected value of the data (e.g., $p_i$ in the Binomial case) with the linear model $\beta_0 + \beta_1 x_i$. For example, the usual link functions for Binomial and Poisson data are the log odds and the log, respectively. These are the defaults in S-PLUS. Alternatives (such as Probit) may be specified.