**Statistics 601**, Spring 2018

**Assignment 2**

**Instructors Comments**

The essence of this issue is whether or not the reduced model given in the assignment is the correct model to use for a likelihood ratio test of difference in binomial parameters. The reduced model given is **not** the correct reduced model for this test. However, if calculation of maximum likelihood estimates and likelihood ratio tests are calculated with no constants included in the log likelihoods, the the values computed from this incorrect model will be the same as those computed using the correct reduced model for the problem, which makes it more difficult to see that there is a difference. The correct full and reduced models for this problem are:

- Full Model: Two binomial distributions,

$$Y_1 \sim \text{Binom}(p_1, n_1) \quad Y_2 \sim \text{Binom}(p_2, n_2),$$

  where $n_1$ and $n_2$ are given.

- Reduced Model: Two binomial distributions with equal parameters,

$$Y_1 \sim \text{Binom}(p, n_1) \quad Y_2 \sim \text{Binom}(p, n_2),$$

1. The full model log likelihood is

$$\begin{aligned} \ell(p_1, p_2) &= y_1 \log(p_1) + (n_1 - y_1)\log(p_1) + \log(n_1) - \log(y_1) - \log(n_1 - y_1) + y_2 \log(p_2) \\ &+ (n_2 - y_2)\log(1 - p_2) + \log(n_2) - \log(y_2) - \log(n_2 - y_2). \end{aligned} \tag{1}$$

2. Maximum likelihood estimates for the full model are

$$\begin{aligned} \hat{p}_1 &= \frac{y_1}{n_1} \\ \hat{p}_2 &= \frac{y_2}{n_2} \end{aligned} \tag{2}$$

3. The log likelihood that corresponds to the incorrect reduced model given in the assignment is

$$\begin{aligned} \ell(p) &= (y_1 + y_2)\log(p) + (n_1 + n_2 - y_1 - y_2)\log(1-p) + \log(n_1+n_2) - \log(y_1+y_2) \\ &- \log(n_1 + n_2 - y_1 - y_2). \end{aligned} \tag{3}$$

4. The log likelihood that corresponds to the correct reduced mdoel given above is

$$\begin{aligned} \ell(p) &= y_1\log(p) + (n_1 - y_1)\log(p) + \log(n_1) - \log(y_1) - \log(n_1 - y_1) + y_2\log(p) \\ &+ (n_2 - y_2)\log(1-p) + \log(n_2) - \log(y_2) - \log(n_2 - y_2) \\ &= (y_1 + y_2)\log(p) + (n_1 + n_2 - y_1 - y_2)\log(1-p) + \log(n_1) - \log(y_1) \\ &- \log(n_1 - y_1) + \log(n_2) - \log(y_2) - \log(n_2 - y_2). \end{aligned} \tag{4}$$

5. Maximizing the incorrect reduced model log likelihood (3) gives the same maximum likelihood estimate of $p$ as does maximizing the correct reduced model log likelihood (4), namely,
$$\hat{p} = \frac{y_1 + y_2}{n_1 + n_2}.$$

6. Dropping constants (terms involving only $n_1$, $n_2$, $y_1$, and/or $y_2$), the likelihood ratio test statistic is the same regardless of whether it is calculated under the incorrect reduced model or the correct reduced model,

$$\begin{aligned} T &= -2[(y_1 + y_2)\log(\hat{p}) - (n_1 + n_2 - y_1 - y_2)\log(1-\hat{p}) - y_1\log(\hat{p}_1) \\ &- y_2\log(\hat{p}_2) + (n_1 - y_1)\log(1-\hat{p}_1) + (n_2 - y_2)\log(1-\hat{p}_2)]. \end{aligned} \tag{5}$$

7. Including constants, the likelihood ratio test statistic for the correct reduced model still gives (5). But the incorrect reduced model produces

$$T = -2\left[(y_1 + y_2)\log(\hat{p}) - (n_1 + n_2 - y_1 - y_2)\log(1-\hat{p}) - y_1\log(\hat{p}_1)\right.$$

$$
\begin{aligned}
- &\quad y_2 \log(\hat{p}_2) + (n_1 - y_1)\log(1-\hat{p}_1) + (n_2 - y_2)\log(1-\hat{p}_2) \\
+ &\quad \log\{(n_1+n_2)!\} - \log\{(y_1+y_2)!\} - \log\{(n_1+n_2-y_1-y_2)!\} \\
- &\quad \log(n_1!) + \log(y_1!) + \log\{(n_1-y_1)!\} - \log(n_2!) + \log(y_2!) + \log\{(n_2-y_2)!\}\big] \, .
\end{aligned} \tag{6}
$$

You will find that expression (6) typically returns a negative number! Because of this, it is clear that the reduced model given in the assignment is not the correct reduced model to test equality of binomial parameters.

8. The parameter spaces of the full and reduced models are the same, regardless of whether the correct or incorrect reduced model is used. And, these spaces are indeed nested. So, there is something more needed than nested parameter spaces to justify a likelihood ratio test. The assignment mentioned page 375 in Casella and Berger. If one looks carefully at that presentation, it will be seen that there is only one likelihood presented, not two. That is, there are not two data models, there is only one. What we call the "reduced model" in a likelihood ratio test is not a different model, really. It is the model with a restriction placed on the parameter. Look again at the assignment. There, you will see two models presented, not one. That is, in using the incorrect reduced model we have actually changed the data model. That means we cannot take the full model, place a restriction on the parameters, and arrive at the reduced model.