Chapter 2 Part 2

04/23/2021

The Case Study Data

- I posted R code to generate our full table of data.

Likelihood Sex Unconditional Model full Data

$$
\begin{aligned}
\mathrm{Lik}(p_B) &= p_B^{930} p_G^{951} p_{BB}^{582} \cdots p_{BBG}^{177} \cdots p_{GGG}^{159} \\
&= p_B^{930+2*582+666+666+\cdots+182} (1-p_B)^{951+666+666+2*530+\cdots+3*159} \\
&= p_B^{5416} (1-p_B)^{5256}
\end{aligned}
$$

MLE of this Likelihood

$$
\begin{aligned}
\hat{p_B} &= \frac{nBoys}{nBoys + nGirls} \\
&= \frac{5416}{5416 + 5256} \\
&= 0.507
\end{aligned}
$$

Likelihood Sex Conditional Model full Data

$$\begin{aligned}
&\mathrm{Lik}(p_{B|N}, p_{B|\mathrm{BBias}}, p_{B|\mathrm{GBias}}) \\
&\quad = [p_{B|N}^{930}(1 - p_{B|N})^{951}(p_{B|N}p_{B|\mathrm{BBias}})^{582} \cdots \\
&\quad ((1 - p_{B|N})(1 - p_{B|\mathrm{GBias}})(1 - p_{B|\mathrm{GBias}}))^{159}] \\
&= [p_{B|N}^{3161}(1-p_{B|N})^{3119} p_{B|\mathrm{BBias}}^{1131}(1-p_{B|\mathrm{BBias}})^{1164} p_{B|\mathrm{GBias}}^{1124}(1-p_{B|\mathrm{GBias}})^{973}]
\end{aligned}$$

Loglikelihood Sex Conditional Model full Data

$$
\begin{aligned}
&\log(\mathrm{Lik}(p_{B|N}, p_{B|\mathrm{BBias}}, p_{B|\mathrm{GBias}})) = \\
&\quad 3161 \log(p_{B|N}) + 3119 \log(1 - p_{B|N}) + 1131 \log(p_{B|\mathrm{BBias}}) \\
&+1164 \log(1-p_{B|\mathrm{BBias}})+1124 \log(p_{B|\mathrm{GBias}})+973 \log(1-p_{B|\mathrm{GBias}})
\end{aligned}
\tag{1}
$$

Resulting MLE's

$$\hat{p}_{B|N} = \frac{3161}{3161 + 3119} = 0.5033$$

$$\hat{p}_{B|\text{B Bias}} = \frac{1131}{1131 + 1164} = 0.4928$$

$$\hat{p}_{B|\text{G Bias}} = \frac{1124}{1124 + 973} = 0.5360$$

Nested models

- Likelihoods are not only useful for fitting models, but they are also useful when comparing models.
- If the parameters for a reduced model are a subset of parameters for a larger model, we say the models are **nested** and the difference between their likelihoods can be incorporated into a statistical test to help judge the benefit of including additional parameters.
- Another way in which to think of nesting is to consider whether parameters in the larger model can be equated to obtain the simpler model or whether some parameters in the larger model can be set to constants.
- Since $p_{B|\text{BBias}}, p_{B|N}$ and $p_{B|\text{GBias}}$ in the Sex Conditional Model can be set to p_B to obtain the Sex Unconditional Model, we can say the models are nested.

Nested models

- If the parameters are not nested, comparing models with the likelihood can still be useful but will take a different form.
- We'll see that the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are functions of the log-likelihood that can be used to compare models even when the models are not nested.
- Either way we see that this notion of likelihood is pretty useful.

Hypotheses

$H_0 : p_{B|N} = p_{B|\mathrm{BBias}} = p_{B|\mathrm{GBias}} = p_B$ (Sex Unconditional Model) The probability of a boy does not depend on the prior family composition.

H_A : At least one parameter from $p_{B|N}, p_{B|\mathrm{BBias}}, p_{B|\mathrm{GBias}}$ differs from the others. (Sex Conditional Model) The probability of a boy does depend on the prior family composition.

Testing This Hypothesis

- We start by comparing the likelihoods or, equivalently, the log-likelihoods of each model at their maxima.
- To do so, we use the log-likelihoods to determine the MLEs, and then replace the parameters in the log-likelihood with their MLEs, thereby finding the maximum value for the log-likelihood of each model.
- Here we will refer to the first model, the Sex Unconditional Model, as the **reduced model**, noting that it has only a single parameter, p_B.
- The more complex model, the Sex Conditional Model, has three parameters and is referred to here as the **larger (full) model**. We'll use the MLEs derived earlier in Section.

Testing This Hypothesis - Smaller (Reduced) Model

- The maximum of the log-likelihood for the reduced model can be found by replacing p_B in the log-likelihood with the MLE of p_B, 0.5075.

$$
\begin{aligned}
\mathrm{Lik}(p_B) &= p_B^{930} p_G^{951} p_{BB}^{582} \cdots p_{BBG}^{177} \cdots p_{GGG}^{159} \\
\log(\mathrm{Lik}(0.5075)) &= 5416 \log(.5075) + 5256 \log(1 - .5075) \\
&= -7396.067
\end{aligned}
$$

Testing This Hypothesis - Full Model

The maximum of the log-likelihood for the larger model can be found by replacing $p_{B|N}, p_{B|\mathrm{BBias}}, p_{B|\mathrm{GBias}}$ in the log-likelihood with 0.5033, 0.4928, and 0.5360, respectively (the MLS's)

$$\begin{aligned}
&\log(\mathrm{Lik}(p_{B|N}, p_{B|\mathrm{BBias}}, p_{B|\mathrm{GBias}})) = \\
&\quad 3161 \log(p_{B|N}) + 3119 \log(1 - p_{B|N}) + 1131 \log(p_{B|\mathrm{BBias}}) \\
&+1164 \log(1-p_{B|\mathrm{BBias}})+1124 \log(p_{B|\mathrm{GBias}})+973 \log(1-p_{B|\mathrm{GBias}})
\end{aligned} \tag{2}$$

$$\begin{aligned}
\log(\mathrm{Lik}(0.5033, 0.4928, 0.5360)) &= 3161 \log(.5033) + 3119 \log(1 - .5033) \\
&+ 1131 \log(.4928) + 1164 \log(1 - .4928) \\
&+ 1124 \log(.5360) + 973 \log(1 - .5360) \\
&= -7391.448
\end{aligned}$$

Comparing log-likelihoods

- The maximum log-likelihood for the larger model is indeed larger (less negative).
- The maximum log-likelihood for the larger model is guaranteed to be at least as large as the maximum log-likelihood for the reduced model, so we'll be interested in whether this observed difference in maximum log-likelihoods, -7391.448 -(-7396.067) $=$ 4.619, is significant.

A Test we Can Use

- When the reduced model is the true model, twice the difference of the maximum log-likelihoods follows a χ^2 distribution with the degrees of freedom equal to the difference in the number of parameters between the two models.
- A difference of the maximum log-likelihoods can also be looked at as the log of the ratio of the likelihoods and for that reason the test is referred to as the **Likelihood Ratio Test (LRT)**.

Our test statistic is

$$\begin{aligned}
\mathrm{LRT} &= 2[\max(\log(\mathrm{Lik}(\text{larger model}))) - \max(\log(\mathrm{Lik}(\text{reduced model})))] \\
&= 2\log\left(\frac{\max(\mathrm{Lik}(\text{larger model}))}{\max(\mathrm{Lik}(\text{reduced model}))}\right)
\end{aligned}$$

Summary

- Intuitively, when the likelihood for the larger model is much greater than it is for the reduced model, we have evidence that the larger model is more closely aligned with the observed data.
- This isn't really a fair comparison on the face of it. We need to account for the fact that more parameters were estimated and used for the larger model.
- That is accomplished by taking into account the degrees of freedom for the χ^2 distribution. The expected value of the χ^2 distribution is its degrees of freedom.

Summary

- Thus when the difference in the number of parameters is large, the test statistic will need to be much larger to convince us that it is not simply chance variation with two identical models.
- Here, under the reduced model we'd expect our test statistic to be 2, when in fact it is over 9. The evidence favors our larger model. More precisely, the test statistic is $2(-7391.448 + 7396.073) = 9.238$ ($p = .0099$), where the p-value is the probability of obtaining a value above 9.238 from a χ^2 distribution with 2 degrees of freedom.

Summary

- We have convincing evidence that the Sex Conditional Model provides a significant improvement over the Sex Unconditional Model.
- However, keep in mind that our point estimates for a probability of a boy were not what we had expected for "sex runs in families."
- It may be that this discrepancy stems from behavioral aspects of family formation.

Note: You may notice that the LRT is similar in spirit to the extra-sum-of-squares F-test used in linear regression. Recall that the extra-sum-of-squares F-test involves comparing two nested models. When the smaller model is true, the F-ratio follows an F-distribution which on average is 1.0. A large, unusual F-ratio provides evidence that the larger model provides a significant improvement.

Non-nested Models

- How does the waiting for a boy model compare to the waiting for a girl model?
- Thus far we've seen how nested models can be compared. But these two models are not nested since one is not simply a reduced version of the other.
- Two measures referred to as information criteria, AIC and BIC, are useful when comparing non-nested models. Each measure can be calculated for a model using a function of the model's maximum log-likelihood. You can find the log-likelihood in the output from most modeling software packages.

Non-nested Models

- $\mathrm{AIC} = -2(\text{maximum log-likelihood }) + 2p$, where p represents the number of parameters in the fitted model. AIC stands for Akaike Information Criterion. Because smaller AICs imply better models, we can think of the second term as a penalty for model complexity—the more variables we use, the larger the AIC.
- $\mathrm{BIC} = -2(\text{maximum log-likelihood }) + p\log(n)$, where p is the number of parameters and n is the number of observations. BIC stands for Bayesian Information Criterion, also known as Schwarz's Bayesian criterion (SBC). Here we see that the penalty for the BIC differs from the AIC, where the log of the number of observations places a greater penalty on each extra predictor, especially for large data sets.

Example

- So which explanation of the data seems more plausible—waiting for a boy or waiting for a girl?
- These models are not nested (i.e., one is not a simplified version of the other), so it is not correct to perform a Likelihood Ratio Test, but we can legitimately compare these models using information criteria.
- We will do this in R
- Note: We skipped the notion of "stopping rules" and as such we have no computed these log likelihoods ourselves
- Stopping rule here is that couples may keep having kids until they have at least 1 of specific sex
- We can still use this example to see how we can compare using criteria

Comparing Criteria

- Smaller AIC and BIC are preferred, so here the Waiting for a Boy Model is judged superior to the Waiting for a Girl Model, suggesting that couples waiting for a boy is a better explanation of the data than waiting for a girl.
- However, for either boys and girls, couples do not stop more frequently after the first occurrence.
- Other stopping rule models are possible. Another model could be that couples wait to stop until they have both a boy and a girl.

Summary of Model Building

- Using a Likelihood Ratio Test, we found statistical evidence that the Sex Conditional Model (Sex Bias) is preferred to the Sex Unconditional Models.
- However, the parameter estimates were not what we expected if we believe that sex runs in families.
- Quite to the contrary, the results suggested that if there were more of one sex in a family, the next child is likely to be of the other sex.
- The results may support the idea that sex composition tends to "even out" over time.

Summary of Model Building

- Using AICs and BICs to compare the non-nested models of waiting for a boy or waiting for a girl, we found that the model specifying stopping for a first boy was superior to the model for stopping for the first girl.
- Again, neither model suggested that couples were *more* likely to stop after the first male or female, rather it appeared just the opposite—couples were *less* likely to be stopping after the first boy or first girl.

Likelihood-Based Methods

- With likelihood methods, we are no longer restricted to independent, identically distributed normal responses (iidN).
- Likelihood methods can accommodate non-normal responses and correlated data.

Linear Regression

- Models that in the past you would fit using ordinary least squares can also be fit using the principle of maximum likelihood.
- It is pleasing to discover that under the right assumptions the maximum likelihood estimates (MLEs) for the intercept and slope in a linear regression are identical to ordinary least squares estimators (OLS) despite the fact that they are obtained in quite different ways.

MLEs are Great

- Beyond the intuitively appealing aspects of MLEs, they also have some very desirable statistical properties.
- You learn more about these features in a statistical theory course.
- Here we briefly summarize the highlights
  - MLEs are *consistent*; i.e., MLEs converge in probability to the true value of the parameter as the sample size increases.
  - MLEs are *asymptotically normal*; as the sample size increases, the distribution of MLEs is closer to normal.
  - MLEs are *efficient* because no consistent estimator has a lower mean squared error. Of all the estimators that produce unbiased estimates of the true parameter value, no estimator will have a smaller mean square error than the MLE.

MLE's Don't Always Exist

- While likelihoods are powerful and flexible, there are times when likelihood-based methods fail:
  - either MLEs do not exist,
  - likelihoods cannot be written down,
  - MLEs cannot be written explicitly.

It is also worth noting that other approaches to the likelihood, such as bootstrapping, can be employed.

Likelihoods and This Course

@Rodgers2001 noted that

> *Many factors have been identified that can potentially affect the human sex ratio at birth. A 1972 paper by Michael Teitelbaum accounted for around 30 such influences, including drinking water, coital rates, parental age, parental socioeconomic status, birth order, and even some societal-level influences like wars and environmental pathogens.*

Likelihoods and This Course

- We ignored these complicating factors to intentionally keep likelihood simple
- Likelihoods answer the sensible question of how likely you are to see your data in different settings.
- When the likelihood is simple as in this chapter, you can roughly determine an MLE by looking at a graph or you can be a little more precise by using calculus or, most conveniently, software.
- As we progress throughout the course, the likelihoods will become more complex and numerical methods may be required to obtain MLEs, yet the concept of an MLE will remain the same.
- Likelihoods will show up in parameter estimation, model performance assessment, and model comparisons.

Likelihoods and Covariates

- One of the reasons many of the likelihoods will become complex is because of covariates.
- Here we estimated probabilities of having a boy in different settings, but we did not use any specific information about families other than sex composition.
- The problems in the remainder of the book will typically employ covariates.
- For example, suppose we had information on paternal age for each family. Consider the Sex Unconditional Model, and let

$$p_B = \frac{e^{\beta_0 + \beta_1 (\text{parental age})}}{1 + e^{\beta_0 + \beta_1 (\text{parental age})}}.$$

(We will give a good reason for this crazy-looking expression for p_B in later chapters.)

Likelihoods and Covariates

- The next step would be to replace p_B in the likelihood, $\text{Lik}(p_B)$, with the complicated expression for p_B.
- The result would be a function of β_0 and β_1. We could then use calculus to find the MLEs for β_0 and β_1.

Likelihoods and Correlated Data

- Another compelling reason for likelihoods occurs when we encounter correlated data.
- For example, models with conditional probabilities do not conform to the *independence assumption*.
- Which of our models violated the independence assumption between parameters?
- We'll see that likelihoods can be useful when the data has structure such as multilevel that induces a correlation. A good portion of the book addresses this.