STATS305B – Lecture 3

Jonathon Taylor
Scribed by Michael Howes

10/03/22

Contents

1 Associations in ordinal data

Last lecture we defined the odds ratio

$$\theta = \frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}},$$

which measured association in a 2×2 table. For $I \times J$ tables, we also defined the local odds ratios

$$\theta_{ij} = \frac{\pi_{ij}\pi_{i+1,j+1}}{\pi_{i,j+1}\pi_{i+1,j}},$$

for $i = 1, \ldots, I-1$ and $j = 1, \ldots, J-1$. The local odds ratios are a collection of $(I-1)(J-1)$ parameters that describe the associations in an $I \times J$ table. If X or Y are ordinal variables rather than categorical variables, then we can use fewer parameters to describe the association. Consider the following table:

	Career satisfaction		
Age	1	2	3
< 30	34	53	88
$30 - 50$	80	174	304
> 50	29	75	172

In this example both X and Y are ordinal, and we would like to know if a higher X (age) results in a higher Y (career satisfaction). To answer this question we can use a single parameter of interested instead of analyzing the four local odd ratios. This is done by studying the probabilities of concordance and discordance.

Note that we can put a partial order on the values (i, j) where $(i, j) \succ (h, k)$ if $i < h$ and $j < k$. In the previous table we would have

$$(> 50, 3) \succ (30 - 50, 2), (30 - 50, 1), (< 30, 2), (< 30, 1).$$

We would not have $(> 50, 3) \succ (> 50, 2)$ or $(> 50, 3) \succ (< 30, 3)$ since we do not allow for either X or Y to tie. A pair of observations (X_1, Y_1) and (X_2, Y_2) are said to be *concordant* if $(X_1, Y_1) \succ (X_2, Y_2)$ or $(X_2, Y_2) \succ (X_1, Y_1)$. In our example, a concordant pair is a pair of individuals where one individual is strictly older, and the older individual has strictly higher career satisfaction. A pair (X_1, Y_1) and (X_2, Y_2) are *discordant* if $(X_1, Y_2) \succ (X_2, Y_1)$ or $(X_2, Y_1) \succ (X_1, Y_1)$. Thus, a discordant pair is again a pair where one individual is strictly older, but now the older individual has strictly lower career satisfaction. We can then define

$$\pi_c = \mathbb{P}(\text{drawing a concordant pair}) = 2 \sum_{ij} \pi_{ij} \sum_{h>i, k>j} \pi_{hk},$$

and

$$\pi_d = \mathbb{P}(\text{drawing a discordant pair}) = 2 \sum_{ij} \pi_{ij} \sum_{h>i, k<j} \pi_{hk}.$$

The parameter

$$\gamma = \frac{\pi_c - \pi_d}{\pi_c + \pi_d},$$

is a measure of the ordinal association between X and Y. Under the null that there is no ordinal association, $\gamma = 0$. We can estimate γ by

$$\widehat{\gamma} = \frac{C - D}{C + D},$$

where

$$C = 2 \sum_{ij} n_{ij} \sum_{h>i, k>j} n_{hk} \quad \text{and} \quad D = 2 \sum_{ij} n_{ij} \sum_{h>i, k<j} n_{hk}.$$

In our example,

$$\frac{1}{2} C = 34 \cdot (174 + 304 + 75 + 172) + 53 \cdot (304 + 172) + 80 \cdot (75 + 172) + 174 \cdot 172 = 99,568,$$

and

$$\frac{1}{2} D = 88 \cdot (80 + 174 + 29 + 75) + 53 \cdot (80 + 29) + 304 \cdot (29 + 75) + 174 \cdot 29 = 73,943.$$

Thus,

$$\widehat{\gamma} = \frac{99568 - 73943}{99568 + 73943} \approx 0.178.$$

Thus, $\widehat{\gamma} \neq 0$ but is this result statistically significant? To answer this we need to know the (approximate) distribution of $\widehat{\gamma}$ under H_0. This leads us to inference for 2-way tables. We will first talk about inference for the odds-ratio, and then we'll return to estimating the distribution of $\widehat{\gamma}$.

2 Inference for 2-way tables (Agresti, Ch 3)

2.1 Wald confidence intervals

Consider the following table documenting car crashes

	Injury (Y)		
Seatbelts (X)	Fatal	Non-Fatal	$\mathbb{P}(\text{Fatal}\|X)$
No	54	10 325	$\approx 0.5\%$
Yes	25	51790	$\approx 0.05\%$.

We would thus estimate the relative risk to be around 10. Since the rare-disease hypothesis seems to hold, we would estimate the odds ratio is also around 10. Our estimator for the odds ratio is

$$\widehat{\theta} = \frac{n_{11}n_{22}}{n_{12}n_{21}} = \frac{54 \times 51790}{25 \times 10325} \approx 10.$$

We wish to understand the standard error of this estimate. To do this, we will consider a number of different models

2.1.1 Poisson sampling

Suppose we have the model $N_{ij} \overset{\text{Indep}}{\sim} \mathsf{Poisson}(\lambda_{ij})$. This model corresponds to when we record all the crashes in a particular place and period. Suppose we wish to test the null $H_0 : X \perp\!\!\!\perp Y$ which is equivalent to $\theta = 1$ where θ is the odds ratio. We know that $\widehat{\theta} \approx 10$, but we wish to calculate the variance of $\widehat{\theta}$. Since $\widehat{\theta} = \frac{N_{11}N_{22}}{N_{12}N_{21}}$, it is easier to work with

$$\log(\widehat{\theta}) = \log(N_{11}) + \log(N_{22}) - \log(N_{12}) - \log(N_{21}).$$

By the independence assumption we have

$$\text{Var}(\log(\widehat{\theta})) = \sum_{i,j=1}^{2} \text{Var}(\log(N_{ij})).$$

To approximate $\text{Var}(\log(N_{ij}))$ we will use the *delta-method*. The delta-method states

$$\text{Var}(f(\widehat{\theta})) \approx \text{Var}(f'(\theta_0)(\widehat{\theta} - \theta_0)) = f'(\theta_0)^2 \,\text{Var}(\widehat{\theta}).$$

Thus, $\text{Var}(\log(N_{ij})) \approx \frac{\text{Var}(N_{ij})}{\lambda_{ij}^2} = \frac{1}{\lambda_{ij}}$. And hence

$$\text{Var}(\log(\widehat{\theta})) \approx \sum_{i,j=1}^{2} \frac{1}{\lambda_{ij}}.$$

We do not know λ_{ij}, but we can estimate λ_{ij} with N_{ij} thus our 95% confidence interval for $\log(\theta)$ is

$$\log(\widehat{\theta}) \pm 1.96 \sqrt{\sum_{i,j=1}^{2} \frac{1}{N_{i,j}}}.$$

By applying the exponential function to the end points we can get a confidence interval for θ. In the seat belt example, the 95% confidence interval for θ is

$$CI = [6.74, 17.42].$$

Thus, with 95% confidence we can reject the null $\theta = 1$.

2.1.2 Multinomial sampling

Instead of using Poisson random variables, we could model our sample using a multinomial distribution

$$(N_{11}, N_{12}, N_{21}, N_{22}) \sim \mathsf{Multinomial}(n, (\pi_{11}, \pi_{12}, \pi_{21}, \pi_{22})).$$

This model would apply if we decided in advance that we would record a random sample of size $n = 60,000$. We again wish to approximate $\text{Var}(\log(\widehat{\theta}))$. Note that

$$\log(\widehat{\theta}) = \log(\widehat{\pi}_{11}) + \log(\widehat{\pi}_{22}) - \log(\widehat{\pi}_{12}) - \log(\widehat{\pi}_{21}) = f(\widehat{\pi}_{11}, \widehat{\pi}_{22}, \widehat{\pi}_{12}, \widehat{\pi}_{12}).$$

The estimates $\widehat{\pi}_{ij}$ are correlated, and so we have to use the multidimensional version of the delta-method:

$$\text{Var}(f(\widehat{\pi})) \approx \nabla f(\pi)^T \, \text{Var}(\widehat{\pi}) \nabla f(\pi).$$

For us, $\nabla f(\pi) = \left[\frac{1}{\pi_{11}}, \frac{1}{\pi_{22}}, -\frac{1}{\pi_{12}}, -\frac{1}{\pi_{21}}\right]^T$ and

$$\text{Var}(\widehat{\pi}) = \frac{1}{n}\left(\text{diag}(\pi) - \pi\pi^T\right).$$

Note that $\text{diag}(\pi)\nabla f(\pi) = [1, 1, -1, -1]^T$ and $\pi^T \nabla f(\pi)) = 0$. Thus,

$$\text{Var}(\log(\widehat{\theta})) \approx \frac{1}{n}\left(f(\pi)^T \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}\right) = \frac{1}{n\pi_{11}} + \frac{1}{n\pi_{22}} + \frac{1}{n\pi_{12}} + \frac{1}{n\pi_{21}}.$$

We do not know π_{ij},but we can estimate $n\pi_{ij}$ with n_{ij}. This gives us the same variance estimate as the Poisson sampling

$$\widehat{\text{Var}}(\log(\widehat{\theta})) = \sum_{i,j=1}^{2} \frac{1}{N_{ij}}.$$

2.1.3 Independent binomial rows

Suppose finally that we have the model

$$N_{11} \sim \mathsf{Binomal}(n_1, \pi_1) \quad \text{and } N_{21} \sim \mathsf{Binomal}(n_2, \pi_2),$$

and $N_{11} \perp\!\!\!\perp N_{21}$. In this model we take a sample of size n_1 of crashes where seatbelts were worn and a sample of size n_2 where seatbelts were not worn. For clinical trials, this type of model is natural. In this case our estimate of θ is

$$\widehat{\theta} = \frac{N_{11}(n_2 - N_{21})}{(n_1 - N_{11})N_{21}} = \frac{\widehat{\pi}_1(1 - \widehat{\pi}_2)}{(1 - \widehat{\pi}_1)\widehat{\pi}_2},$$

and

$$\log(\widehat{\theta}) = \log(\widehat{\pi}_1) + \log(1 - \widehat{\pi}_2) - \log(1 - \widehat{\pi}_1) - \log(\widehat{\pi}_2).$$

By independence,

$$\text{Var}(\log(\widehat{\theta})) = \text{Var}(\log(\widehat{\pi}_1) - \log(1 - \widehat{\pi}_1)) + \text{Var}(\log(\widehat{\pi}_2) - \log(1 - \widehat{\pi}_2)).$$

Let $f(\pi_i) = \log(\pi_i) - \log(1 - \pi_i)$. Then $f'(\pi_i) = \frac{1}{\pi_i} + \frac{1}{1-\pi_i} = \frac{1}{\pi_i(1-\pi_i)}$. Also, $\text{Var}(\widehat{\pi}_i) = \frac{\pi_i(1-\pi_i)}{n_i}$. Thus, by the delta-method

$$\text{Var}(\log(\widehat{\pi}_1) - \log(1 - \widehat{\pi}_1)) \approx \frac{1}{(\pi_i(1 - \pi_i))^2} \cdot \frac{\pi_i(1 - \pi_i)}{n_i} = \frac{1}{n_i \pi_i (1 - \pi_i)}.$$

Thus,

$$\mathrm{Var}(\log(\widehat{\theta})) \approx \frac{1}{n_1\pi_1(1-\pi_1)} + \frac{1}{n_2\pi_2(1-\pi_2)}.$$

We estimate π_i with $\frac{N_{i1}}{n_i}$. This gives,

$$\begin{aligned}\widehat{\mathrm{Var}}(\log(\widehat{\theta})) &= \frac{1}{N_{11}(1-N_{11}/n_1)} + \frac{1}{N_{21}(1-N_{21}/n_2)} \\ &= \frac{n_1}{N_{11}N_{12}} + \frac{n_2}{N_{21}N_{22}} \\ &= \frac{N_{11}+N_{12}}{N_{11}N_{12}} + \frac{N_{21}+N_{22}}{N_{21}N_{22}} \\ &= \sum_{i,j=1}^{2} \frac{1}{N_{i,j}}.\end{aligned}$$

Thus, we get the same estimate as before. Each of these calculations are examples of "Wald confidence intervals/significance tests". In each case we estimate the standard errors by plugging in the MLE. This is different to likelihood ratio tests and score tests. For likelihood ratio tests and score tests we estimate the standard errors by plugging the MLE under the null.

2.2 Profile likelihood

Suppose we are interested in a parameter $\theta = f(\pi)$, and we wish to do maximum likelihood estimation. There may be multiple parameters π that give the same value of θ. This means that we have to work with the *profile likelihood*. For a fixed θ_0, consider the set

$$A(\theta_0) = \{\pi : f(\pi) = \theta_0\}.$$

And define

$$\psi(\theta_0|Y) = \operatorname*{argmax}_{\pi \in A(\theta_0)} L(\pi|Y),$$

where Y is our data and L is the likelihood. The value $\psi(\theta_0|Y)$ is the MLE under the null $H_0 : \theta = \theta_0$. The *profile likelihood* is the function

$$\theta \mapsto L(\psi(\theta|Y)|Y) = \max_{\pi : f(\pi)=\theta} L(\pi|Y).$$

Maximizing the profile likelihood is equivalent to maximizing the original likelihood in that

$$\widehat{\theta} = \operatorname*{argmax}_{\theta} L(\psi(\theta|Y)|Y) = f(\operatorname*{argmax}_{\pi} L(\theta|Y)) = f(\widehat{\pi}).$$

We can use the profile likelihood to define confidence intervals for θ. In particular

$$CI = \{\theta : -2\log(L(\psi(\theta|Y)|Y)) + 2\log(L(\widehat{\pi}|Y)) \leq \chi^2_{1,1-\alpha}\},$$

is a $100(1-\alpha)\%$ confidence set for θ.

2.3 Relative risk

Suppose we have a 2×2 table and we wish to estimate the relative risk $= \frac{\pi_1}{\pi_2} = \frac{\mathbb{P}(Y=1|X=1)}{\mathbb{P}(Y=1|X=2)}$. We can use the estimator

$$\widehat{r} = \frac{\widehat{\pi}_1}{\widehat{\pi}_2} = \frac{n_{11}/n_{1+}}{n_{21}/n_{2+}}.$$

Thus,

$$\log(\widehat{r}) = \log(\widehat{\pi}_1) - \log(\widehat{\pi}_2).$$

Suppose we have a clinical trial and so $N_{1i} \overset{\text{Indep}}{\sim} \mathsf{Binomal}(n_i, \pi_i)$. Then

$$\text{Var}(\log(\widehat{r})) = \text{Var}(\log(\widehat{\pi}_1)) + \text{Var}(\log(\widehat{\pi}_2)).$$

By the delta method we would have,

$$\text{Var}(\log(\widehat{\pi}_i)) \approx \frac{1}{\pi_i^2} \frac{\pi_i(1-\pi_i)}{n_i} = \frac{1-\pi_i}{\pi_i n_i}.$$

We can estimate the approximate variance by plugging in the MLE

$$\widehat{\text{Var}}(\log(\widehat{r})) = \frac{1-\widehat{\pi}_1}{N_{11}} + \frac{1-\widehat{\pi}_2}{N_{22}}.$$

The Poisson model can also be used to estimate the standard error of the relative risk and the result using the delta-method is the same.

3 Inference for $I \times J$ tables (Agresti 3.2)

Suppose we have a table

		Y			
		1	2	…	J
X	1				
	2				
	⋮				
	I				

where we have counts $N_{ij} \sim \mathsf{Poisson}(\lambda_{ij})$ and thus $\pi_{ij} = \mathbb{P}(X = i, Y = j) = \frac{\lambda_{ij}}{\sum_{h,k} \lambda_{h,k}}$. Let $\pi_{i+} = \mathbb{P}(X = i) = \sum_{j=1}^{J} \pi_{ij}$ and $\pi_{+j} = \mathbb{P}(Y = j) = \sum_{i=1}^{I} \pi_{ij}$. Suppose we want to test for independence of X and Y. Our null is thus $H_0 : \pi_{ij} = \pi_{i+}\pi_{+j}$ for all i and j and our alternative is that π_{ij} are unconstrained. We can do a likelihood ratio test or Pearson's χ^2 test.

3.1 Likelihood ratio test

Let $L_0 = \max_{\lambda \in H_0} L(\lambda|N)$ and let $L_1 = \max_{\lambda \in H_0 \cup H_1} L(\lambda|N)$. Under the null, the statistic

$$G^2 = -2\log L_0 - (-2\log L_1),$$

has asymptotic distribution $\chi^2_{(I-1)(J-1)}$. This is because $H_0 \cup H_1$ has IJ free parameters and H_0 has $(I-1) + (J-1) + 1$ free parameters. We know that the MLE under $H_0 \cup H_1$ is $\widehat{\pi}_{ij} = \frac{n_{ij}}{N_{++}}$ and $\widehat{\lambda} = N_{++}$ where $\lambda = \sum_{i,j} \lambda_{ij}$. We will see that the MLEs under H_0 are $\pi_{i+} = \frac{N_{i+}}{N_{++}}$, $\pi_{+j} = \frac{N_{+j}}{N_{++}}$ and $\widehat{\sum_{h,k} \lambda_{hk}} = N_{++}$. Thus,

$$-2\log(L_0) = 2\left[\sum_{ij} -N_{ij}\log(\widehat{\lambda}_{ij}) + \widehat{\lambda}_{ij}\right]$$

$$-2\log(L_1) = 2\left[\sum_{ij} -N_{ij}\log(N_{ij}) + N_{ij}\right]$$

Since $\sum_{ij} N_{ij} = \sum_{ij} \widehat{\lambda}_{ij}$, we have

$$G^2 = 2 \sum_{ij} N_{ij} \log \left(\frac{N_{ij}}{\widehat{\lambda}_{ij}} \right).$$

We now verify our MLE estimates under H_0. We begin by re-parametrizing, if $(\lambda_{ij}) \in H_0$, then we have

$$\lambda_{ij} = e^{\mu} \pi_{i+} \pi_{+j} = e^{\mu} e^{\alpha_i} e^{\beta_j}.$$

Thus, we will optimize μ, α_i and β_j. Note that

$$-\log(L(\mu, \alpha, \beta | N)) = \sum_{ij} (-N_{ij}(\mu + \alpha_i + \beta_j) + e^{\mu + \alpha_i + \beta_j}).$$

Thus,

$$\frac{\partial}{\partial \alpha_i} - \log(L(\mu, \alpha, \beta | N)) = -\sum_{ij} N_{ij} + e^{\mu + \alpha_i + \beta_j}.$$

Setting the partials equal to zero gives

$$e^{\widehat{\alpha}_i} \left(\sum_j e^{\widehat{\beta}_j} \right) e^{\widehat{\mu}} = \sum_{j=1}^{J} N_{ij} = N_{i+}.$$

And similarly for $\widehat{\beta}_j$. One can then check that

$$e^{\widehat{\alpha}_i} = \frac{N_{i+}}{N_{++}}, \quad e^{\widehat{\beta}_j} = \frac{N_{+j}}{N_{++}} \quad \text{and} \quad e^{\widehat{\mu}} = N_{++},$$

solve the first order equations and thus are MLEs.

3.2 Pearson's test

Pearson's χ^2 test is a type of score test with statistic

$$(\widehat{\theta} - \theta_0)^T \operatorname{Var}(\theta_0)^{-1} (\widehat{\theta} - \theta_0).$$

For the test of independence, Pearson's test statistic is

$$X^2 = \sum_{ij} \frac{(N_{ij} - \widehat{\lambda}_{ij})^2}{\widehat{\lambda}_{ij}} = \sum_{ij} \frac{(O_{ij} - E_{ij})^2}{E_{ij}},$$

where $E_{ij} = \widehat{\lambda}_{ij} = \frac{n_{i+} n_{+j}}{n_{++}}$ are the MLE estimates under H_0 (also called the expected values) and $O_{ij} = N_{ij}$ are the observed values. Under H_0, $X^2 \sim \chi^2_{(I-1)(J-1)}$, asymptotically.

4 Bootstrapping

Returning to the statistics $\widehat{\gamma}$ from Section 1. Calculating a standard error for $\widehat{\gamma}$ is harder, but there are r packages that do this. Another option is to bootstrap. To do this we can create a "flattened" version of the contingency table where the number of rows is N_{++} the total number of individuals. And in each row we have a record of each individual's X and Y response. This means there will be exactly N_{ij} rows with $X = i$ and $Y = j$. We can then create bootstrap samples (samples *with* replacement) from the flattened table. We can then create contingency tables for each bootstrapped sample and then calculate $\widehat{\gamma}^*_b$ for each bootstrap sample b. Then we use the empirical distribution of $\{\widehat{\gamma}^*_b\}_{b=1}^B$ to estimate the sampling distribution of $\widehat{\gamma}$. We'll talk more about this next lecture.