STAT 557 Homework 2

Yifan Zhu

October 5, 2017

1. (a) It could be a multinomial distribution $Mult\left(n, \pi_{11}, \pi_{12}, \pi_{21}, \pi_{22}\right)$. $n = 1397$ in this problem.

(b) Let i corresponds to Gun Registration and j corresponds to Death Penalty. i, j take values 1 and 2. 1 for favor and 2 for oppose. Then we have

$$\begin{aligned}\hat{m}_{11} &= 799.50\\ \hat{m}_{12} &= 220.50\\ \hat{m}_{21} &= 295.50\\ \hat{m}_{22} &= 81.50\end{aligned}$$

(c)

$$\begin{aligned}G^2 &= \sum_{i=1}^{2}\sum_{j=1}^{2} y_{ij}\log\left(\frac{y_{ij}}{\hat{m}_{ij}}\right) = 5.32\\ X^2 &= \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{(y_{ij}-\hat{m}_{ij})^2}{\hat{m}_{ij}} = 5.15\end{aligned}$$

The reference distribution is chi-squared distribution with degrees of freedom to be **1**. The p-values for G^2 is **0.021** and for X^2 is **0.023**. The p-values are small and we reject the null hypothesis and conclude that there is sufficient evidence that gun registration opinions is **not** held independently of death penalty opinions.

2. (a)

$$\begin{aligned}\ell(\theta) &= \log n! - \sum_{i=1}^{2}\sum_{j=1}^{2}\log Y_{ij}! + \sum_{i=1}^{2}\sum_{j=1}^{2} Y_{ij}\log(\pi_{ij})\\ &= \log n! - \sum_{i=1}^{2}\sum_{j=1}^{2}\log Y_{ij} + (2Y_{11}+Y_{12}+Y_{21})\log\theta + (2Y_{22}+Y_{12}+Y_{21})\log(1-\theta)\end{aligned}$$

(b)

$$\begin{aligned}\frac{\mathrm{d}\ell}{\mathrm{d}\theta} &= \frac{2Y_{11}+Y_{12}+Y_{21}}{\theta} - \frac{2Y_{22}+Y_{12}+Y_{21}}{1-\theta}\\ \frac{\mathrm{d}^2\ell}{\mathrm{d}\theta^1} &= -\frac{2Y_{11}+Y_{12}+Y_{21}}{\theta^2} - \frac{2Y_{22}+Y_{12}+Y_{21}}{(1-\theta)^2} < 0\end{aligned}$$

Hence $\frac{\mathrm{d}\ell}{\mathrm{d}\theta}\big|_{\theta=\hat{\theta}} = 0 \Rightarrow$

$$\hat{\theta} = \frac{2Y_{11}+Y_{12}+Y_{21}}{2Y_{11}+Y_{12}+Y_{21}+Y_{22}+Y_{12}+Y_{21}} = \frac{2Y_{11}+Y_{21}+Y_{12}}{2n}$$

(c)

$$
\begin{aligned}
\hat{m}_{11,A} &= \frac{(2Y_{11}+Y_{12}+Y_{22})^2}{4n} \\
\hat{m}_{12,A} = \hat{m}_{21} &= \frac{(2Y_{11}+Y_{12}+Y_{21})(2Y_{22}+Y_{12}+Y_{21})}{4n} \\
\hat{m}_{22,A} &= \frac{(2Y_{22}+Y_{12}+Y_{21})^2}{4n}
\end{aligned}
$$

(d)

$$
\begin{aligned}
G^2 &= 2\sum_{i=1}^{2}\sum_{j=1}^{2} y_{ij}\log\left(\frac{y_{ij}}{\hat{m}_{ij,A}}\right) \\
&= 2\left(y_{11}\log\left(\frac{4ny_{11}}{(2y_{11}+y_{12}+y_{21})^2}\right) + y_{12}\log\left(\frac{4ny_{12}}{(2y_{11}+y_{12}+y_{21})(2y_{22}+y_{12}+y_{21})}\right)\right. \\
&\left. + y_{21}\log\left(\frac{4ny_{21}}{(2y_{11}+y_{12}+y_{21})(2y_{22}+y_{12}+y_{21})}\right) + y_{22}\log\left(\frac{4ny_{22}}{(2y_{22}+y_{12}+y_{21})^2}\right)\right)
\end{aligned}
$$

Degrees of freedom is **2**.

(e) Let $\pi_{11} = \theta_1, \pi_{12} = \theta_2$, from $\pi_{i+} = \pi_{+i}$ we have $\pi_{21} = \theta_2, \pi_{22} = \theta_3$. Hence

$$
\ell(\theta_1,\theta_2,\theta_3) = \log n! - \sum_{i=1}^{2}\sum_{j=1}^{2}\log Y_{ij}! + Y_{11}\log\theta_1 + (Y_{12}+Y_{21})\log\theta_2 + Y_{22}\log\theta_3
$$

Maximize

$$
g(\theta_1,\theta_2,\theta_3,\lambda) = \ell(\theta_1,\theta_2,\theta_3) + \lambda(1-\theta_1-2\theta_2-\theta_3)
$$

$$
\begin{aligned}
\frac{\partial g}{\theta_1} &= \frac{Y_{11}}{\theta_1} - \lambda \\
\frac{\partial g}{\theta_2} &= \frac{Y_{12}+Y_{21}}{\theta_2} - 2\lambda \\
\frac{\partial g}{\theta_3} &= \frac{Y_{22}}{\theta_3} - \lambda \\
\frac{\partial g}{\partial\lambda} &= 1-\theta_1-2\theta_2-\theta_3 = 0
\end{aligned}
$$

Then we have

$$
\begin{aligned}
\hat{\theta}_1 &= \frac{y_{11}}{n} \\
\hat{\theta}_2 &= \frac{y_{12}+y_{21}}{2n} \\
\hat{\theta}_3 &= \frac{y_{22}}{n}
\end{aligned}
$$

Then

$$
\begin{aligned}
\hat{m}_{11,B} &= Y_{11} \\
\hat{m}_{22,B} &= Y_{22} \\
\hat{m}_{12,B} = \hat{m}_{21,B} &= \frac{Y_{12}+Y_{21}}{2}
\end{aligned}
$$

$$
\begin{aligned}
G^2 &= 2\sum_{i=1}^{2}\sum_{j=1}^{2} y_{ij}\log\left(\frac{y_{ij}}{\hat{m}_{ij,B}}\right) \\
&= 2\left(y_{12}\log\left(\frac{2y_{12}}{y_{12}+y_{21}}\right) + y_{21}\log\left(\frac{2y_{21}}{y_{12}+y_{21}}\right)\right)
\end{aligned}
$$

Degrees of freedom is **1**.

(f) Model A and Model B don't fit well. But Model B fits significantly better then Model A.

Comparison	d.f.	deviance value	p-value
Model A vs Model B	1	5.966123	0.01458331
Model B vs Model C	1	10.31583	0.00131894
Model A vs Model C	2	16.28195	0.0002913527

3. (a)

$$\hat{m} = \frac{1}{k}\sum_{i=1}^{k} Y_i = 3.87$$

(b) The Pearson statistic is **12.9641**, degrees of freedom is **10** and p-value is **0.2257**. The p-value is large, so we fail to reject the null hypothesis and conclude that the Poisson model fits well here.

Number of Scintillations	Number of Time Intervals	Expected Counts Poisson Model
0	57	54.314425
1	203	210.280962
2	383	407.056532
3	525	525.313114
4	532	508.443876
5	408	393.693084
6	273	254.033683
7	139	140.500553
8	45	67.994348
9	27	29.249273
10	10	11.324000
11 - 14	6	5.760294

(c) Fisher's dispersion index is **2488.918**, degrees of freedom is **2607** and p-value is **0.102**. With large p-value we conclude that the number of scintillations are consistent with i.i.d Poisson(m) model.

(d) Using the exact confidence interval

$$\left[\frac{\chi^2_{2\sum_j Y_j, 0.025}}{2k}, \frac{\chi^2_{2+2\sum_j Y_j, 0.975}}{2k}\right] = [3.796397, 3.947814]$$

(e) The algorithm fails to converge. Because the Poisson model fits well for the data, and if a Negative-Binomial need to approximate a Poisson distribution, we need $\beta = k \to \infty$. Thus it dose not converge.

4. (a)

$$\hat{m} = 3.156627$$

(b) The Pearson statistic is **26.57834**, degrees of freedom is **7** and p-value is **0.0004**. The p-value is small, so we reject the null hypothesis and conclude that the Poisson model does **not** fit well here.

Number of Accidents	**Number of Drivers**	**Expected Counts Poisson Model**
0	15	7.066472
1	32	22.306211
2	26	35.206189
3	29	37.044263
4	22	29.233726
5	19	18.455991
6	9	9.709778
7	8	4.378592
8 - 15	6	2.598738

(c) Fisher's dispersion index is **293.9618**, degrees of freedom is **165** and p-value is **2.647e-09**. With small p-value we conclude that the number of scintillations are **not** consistent with i.i.d Poisson(m) model.

(d) Using the exact confidence interval

$$\left[\frac{\chi^2_{2\sum_j Y_j,0.025}}{2k}, \frac{\chi^2_{2+2\sum_j Y_j,0.975}}{2k}\right] = [2.892102, 3.438843]$$

(e)

$$\hat{\pi} = 0.575, \hat{\beta} = 4.272$$

(f) The Pearson statistic is **3.989601**, degrees of freedom is **8** and p-value is **0.858**. The p-value is large, so we fail to reject the null hypothesis and conclude that the Negative-Binomial model fit well here.

Number of Accidents	**Number of Drivers**	**Expected Counts Poisson Model**
0	15	15.619369
1	32	28.352669
2	26	31.757440
3	29	28.212469
4	22	21.794615
5	19	15.321685
6	9	10.061148
7	8	6.273718
8	3	3.756278
9	1	2.176474
10-15	2	2.674135

(g)

$$\hat{m} = \hat{\beta}\left(\frac{1-\hat{\pi}}{\hat{\pi}}\right) = 3.156626$$

And

$$\frac{\partial m}{\partial \pi} = -\frac{\beta}{\pi^2}, \frac{\partial m}{\partial \beta} = \frac{1-\pi}{\pi}$$

Let $D = \left.\begin{bmatrix} \frac{\partial m}{\partial \pi} & \frac{\partial m}{\partial \beta} \end{bmatrix}^T\right|_{\hat{\pi},\hat{\beta}}$. Then by Delta Method, the variance of $\hat{m}$ is

$$\sigma^2 = D^T \widehat{Cov}(\pi, \beta) D = 0.03306799$$

Then standard error for the 95% confidence interval is

$$\sigma = \sqrt{0.03306799} = 0.1818461$$

The 95% confidence interval is

$$(2.800208, 3.513045)$$

5. (a) It is a retrospective study.

(b) Approximate 95% confidence interval for $\log \alpha$ is

$$\log(2.93) \pm 1.96\sqrt{\frac{1}{67}+\frac{1}{34}+\frac{1}{43}+\frac{1}{64}} \Rightarrow (0.509591, 1.640414)$$

So 95% confidence interval for α is

$$(\exp(0.509591), \exp(1.640414)) = (1.66461, 5.157304)$$

(c) I agree with Vienna's. In the second analysis, the control group and the Hodgkin's Disease group are obvious not independent. So the odds should appear to be similar for two groups.

6. The p-value is 0.6384. With a large p-value we fail to reject the null hypothesis and conclude that the two treatments are equally effective.

7. (a) $G^2 = 9.5142$, degrees of freedom is **3** and p-value is **0.023**.

(b) p-value = 0.014.

(c) $X^2 = 9.5110$, degrees of freedom is **3** and p-value is **0.023**.

(d) p-value = 0.014

(e) The p-value is **0.01338** from 50000 simulation. It is smaller then above.

8. (a) For female, $G^2 = 47.817$, degrees of freedom is **4** and p-value is **1.03e-09**.$X^2 = 50.254$, degrees of freedom is **4** and p-value is **3.20e-10**. With small p-value we conclude that physical and psychological demands are not independent.

For male, $G^2 = 37.509$, degrees of freedom is **4** and p-value is **1.4145e-07**.$X^2 = 35.909$, degrees of freedom is **4** and p-value is **3.0212e-07**. With small p-value we conclude that physical and psychological demands are not independent.

(b) For female, $\hat{\gamma}_1 = 0.2356757$, the standard error $\hat{\sigma}_1$ is **0.03848595** and 95% confidence interval is [0.160, 0.311].

For male, $\hat{\gamma}_2 = -0.12533285$, the standard error $\hat{\sigma}_2$ is **0.03271634** and 95% confidence interval is [-0.189, -0.061].

9. (a)

$$\frac{d\zeta}{d\gamma} = \frac{1}{1-\gamma^2} \Rightarrow Var(\hat{\zeta}) = \frac{Var(\hat{\gamma})}{(1-\hat{\gamma}^2)^2}$$

(b)

(c)

10. (a) For White children, 95% confidence interval for Cohen's Kappa is (-0.24526, -0.0038546). For Black children, 95% confidence interval for Cohen's Kappa is (-0.18027, 0.03269). From the confidence interval we can see that for black children, there seems to be random agreement, and for white children, there is no random agreement.

(b) For white children, 95% confidence interval using bootstrap is (-0.2466, -0.0093). For black children, 95% confidence interval using bootstrap is (-0.1812, 0.0223). Both of them are narrower than the CI's from (a).

(c) Standard Error for difference is $\sqrt{\hat{\sigma}_1^2 + \hat{\sigma}_2^2} = \sqrt{0.061583^2 + 0.054328^2} = 0.006743997$. The difference is $\hat{\kappa}_1 - \hat{\kappa_2} = -0.12456 - (-0.073792) = -0.050768$. Hence the 95% confidence interval is

$$(-0.06398623, -0.03754977)$$

, which does not include 0. So the level of agreement are not the same.

11. (a)

$$\hat{\alpha}_{MH} = 3.00465$$

(b) The 95% confidence interval is $(2.279925, 3.959745)$. We are 95% confident that the common odds ratio is between (2.279925,3.959745).

(c) Breslow-Day statistic is **9.6924**. Degrees of freedom is **2**. p-value is **0.007858**. With low p-value, we reject null hypothesis and conclude that the odds ratios are different.

(d) Exact 95% confidence interval is (2.278093,4.043850).

(e) p-value is **0.0082**. We reject null hypothesis and conclude that the odds ratios are different.

(f) No. Because from several tests we have the same conclusion that the odds ratios are different.