An Outline of Solution to STAT 545 Midterm Exam - 10/15

1. (**30 points**)

   (a) (**15 points**)

   - (**5 points**)
     $OR = 2 * 8/(3 * 3)$ (**3 points**)
     Interpretation: The odds of Y's falling in category 1 at level 1 of X is (2*8/(3*3)) times that at level 2 of X; (**2 points**)
   - (**5 points**)
     Diff in prop $= 2/5 - 3/11$ (**3 points**)
     Interpretation: The proportion of being in category 1 of Y at level 1 of X is (2/5-3/11) more than that at level 2 of X, or the difference in the proportion of being in category 1 of Y between levels 1 and 2 of X is (2/5-3/11); (**2 points**)
   - (**5 points**)
     $RR = (2/5)/(3/11)$ (**3 points**)
     Interpretation: The proportion of being in category 1 of Y at level 1 of X is ((2/5)/(3/11)) times that at level 2 of X; (**2 points**)

   (b) (**15 points**)

   - (**5 points**)
     $\log OR = \log(2 * 8/(3 * 3))$, (**1 point**)
     $SE(\log OR) = \sqrt{1/2 + 1/3 + 1/3 + 1/8}$, (**1 point**)
     $\log OR \pm 1.96 * SE(\log OR)$, (**2 points**)
     $95\% CL : (\exp(LL), \exp(RL))$. (**1 point**)
   - (**5 points**)
     Diff in prop $= 2/5 - 3/11$, (**1 point**)
     $SE(\text{Diff in prop}) = \sqrt{2/5(1 - 2/5)/5 + 3/11(1 - 3/11)/11}$, (**2 points**)
     Diff in prop $\pm\ 1.96 * SE(\text{Diff in prop})$. (**2 points**)
   - (**5 points**)

$\log RR = \log(2/5/(3/11))$, (**1 point**)

$SE(\log RR) = \sqrt{1/2 - 1/5 + 1/3 - 1/11}$, (**1 point**)

$LL = \log(RR) - 1.96 * SE(\log(RR))$, $RL = \log(RR) + 1.96 * SE(\log(RR))$, (**2 points**)

$95\% CL : (\exp LL, \exp RL)$. (**1 point**)

2. (**30 points**)

(a) (**14 points**) Pearson χ^2 statistic:

$$X^2 = \sum_i \sum_j (n_{ij} - \hat{m}_{ij})^2 / \hat{m}_{ij},$$

where $\hat{m}_{ij} = n_{i+} n_{+j} / n$ with $n_{i+} = \sum_k n_{ik}$ and $n_{+j} = \sum_k n_{kj}$, $i = 1, 2$, $j = 1, 2$. (**5 points**) $DF = 1$. (**2 points**)

Likelihood ratio χ^2 statistics:

$$G^2 = 2 \sum_i \sum_j n_{ij} \log(n_{ij} / \hat{m}_{ij}),$$

where $\hat{m}_{ij}$ are defined the same as above. (**5 points**) $Df = 1$. (**2 points**)

(b) (**16 points**) Should not use the test statistics from (a). (**2 points**)

Reason: Pearson and likelihood ratio χ^2 tests are based on large samples, while the data given are from a small sample. The Pearson and likelihood ratio tests may not be valid. (**4 points**)

An alternative: Fisher's exact test. Hypergeometric distribution for n_{11} conditional on both row and column totals:

$$\frac{\begin{pmatrix} n_{1+} \\ n_{11} \end{pmatrix} \begin{pmatrix} n_{2+} \\ n_{+1} - n_{11} \end{pmatrix}}{\begin{pmatrix} n \\ n_{+1} \end{pmatrix}},$$

where $n_{11} \in [m_-, m_+]$ with $m_- = \max(0, n_{1+} + n_{+1} - n)$ and $m_+ = \min(n_{1+}, n_{+1})$. (**6 points**)

P-value:

$$\frac{\begin{pmatrix}4\\1\end{pmatrix}\begin{pmatrix}4\\3\end{pmatrix}}{\begin{pmatrix}8\\4\end{pmatrix}}+\frac{\begin{pmatrix}4\\0\end{pmatrix}\begin{pmatrix}4\\4\end{pmatrix}}{\begin{pmatrix}8\\4\end{pmatrix}}$$

(**4 points**)

3. (**20 points**)

   (a) (**10 points**)

   - (**8 points**) $L(\theta) = \theta^{2n_{11}}\theta^{n_{12}+n_{21}}(1-\theta)^{n_{12}+n_{21}}(1-\theta)^{2n_{22}}$. Solve $L'(\theta) = 0$ for $\theta \Longrightarrow \hat{\theta} = (p_{1+} + p_{+1})/2$.
   - (**2 points**) Show $\hat{\theta}$ actually maximizes the likelihood, e.g., by showing the second derive at $\hat{\theta} < 0$.

   (b) (**10 points**)

   - (**6 points**) $X^2 = \sum_i \sum_j (n_{ij} - \hat{m}_{ij})^2/\hat{m}_{ij}$ with $\hat{m}_{11} = n\hat{\theta}^2, \hat{m}_{12} = \hat{m}_{21} = n\hat{\theta}(1-\hat{\theta})$, and $\hat{m}_{22} = n(1-\hat{\theta})^2$.
   - (**4 points**) $DF = 2$.

4. (**20 points**)

   (a) (**10 points**) Let $\sum_{y=1}^{\infty} r(\theta)\exp(\theta y)/y! = 1$. (**2 points**)

   We have $r(\theta) = \left\{\sum_{y=1}^{\infty} \exp(\theta y)/y!\right\}^{-1} = \left\{\sum_{y=1}^{\infty} (\exp(\theta))^y/y!\right\}^{-1}$ (**3 points**) $= \{\exp(\exp(\theta)) - 1\}^{-1}$. (**5 points**)

   (b) (**10 points**) $f_Y(y;\theta)$ is an exponential-family distribution with $\theta = \theta$, $b(\theta) = \log(\exp(\exp(\theta)) - 1)$, $\phi = a(\phi) = 1$, $c(y,\phi) = \log y!$. (**4 points**)

   The kernel of the log-likelihood function for a sample of i.i.d. observations $Y_1, \ldots, Y_n$ is $\theta \sum_{i=1}^{n} Y_i - n\log(\exp(\exp(\theta)) - 1)$. (**2 points**)

   The likelihood equation therefore is $d\log(\exp(\exp(\theta)) - 1)/d\theta = \bar{Y}$, where $\bar{Y} = \sum_{i=1}^{n} Y_i/n$. (**2 points**)

After simplification, we get

$$\frac{\exp(\theta)}{1 - \exp(-\exp(\theta))} = \bar{Y}. \text{ (\textbf{2 points})}$$