Lecture 1

I. Introduction

A: 1.1

1. Course investigates description and analysis of variables that are \textit{categorical}
   1. \textit{e}., variables that name categories
   2. Ex.,
      a. Healthy/Sick
      b. Young/Old
      c. Strongly Agree/Agree/Indifferent/Disagree/Strongly Disagree
      d. Alabama/Alaska/\ldots/Wyoming
   A: 1.1.2

2. Categories might be ordered or unordered
   a. Careful: can get fooled
      i. Ex., zip codes
   b. If \( X \) categories (numbered from zero)
      a. Here indexed by \( x \)
      i. Lower bound \( i \)
      ii. \( \Phi \) probilities
   c. This distribution depends on the \( \pi \)
   A: 1.1.1
   d. Number of individuals in each category is given by a
   d. Poisson distribution with intensity \( \lambda_k \)
   i. \( P [X_k = x_k] = \exp(-\lambda_k)\lambda_k^{x_k} / x_k! \)
   e. Assume entries are independent
   f. Well known result: \( X_i \sim \mathcal{P}(\lambda_i) \), \( X_j \sim \mathcal{P}(\lambda_j) \) \( \Rightarrow \)
   \( X_i + X_j \sim \mathcal{P}(\lambda_i + \lambda_j) \).
   i. Obvious extension to more summands
   A: 1.2.1-1.2.2

2. Multinomial:
   a. Notation: Quantities with a + for the subscript
      indicate summation over that subscript.
   b. Conditional on observing \( X_+ = x_+ \),
   \( P [X_0 = x_0, \ldots, X_{K-1} = x_{K-1} | X_+ = x_] = \prod_{k=0}^{K-1} \exp(-\lambda_k)\lambda_k^{x_k} / x_k! / \exp(-\lambda_+)\lambda_+^{x_+} / x_+! \) \( \Rightarrow \)
   \( \prod_{k=0}^{K-1} \pi_k x_k! / x_+! \) \( = \prod_{k=0}^{K-1} \pi_k x_k! / x_+! \) \( \Rightarrow \)
   \( \pi_k = \lambda_k / \lambda_+ \)
   c. This distribution depends on the \( \lambda_0, \ldots, \lambda_{K-1} \) only
      through the ratios \( \lambda_k / \lambda_+ \).
   d. Equivalent to taking each Poisson arrives to any
      category, and dividing them among bins according to
      probabilities \( \pi_k \).
   e. Special Case: \( K = 2 \)
   \( X_0/X_+ \sim \mathcal{N}(\pi_0, \pi_1) \).
   B. Conditionality principal:
   1. data arises from random mixture of experiments
   a. Here indexed by \( X_+ \)

II. Inference on a Single Categorical Variable:

A. Distributional Assumptions:

A: 3.3

1. Poisson
   a. Suppose that there are \( K \) categories (numbered from zero)
   b. Let \( X_k \) be the (random) number in category \( k \)
   c. Let \( x_k \) be a potential value for \( X_k \)

2. mixing distribution does not depend on unknown parameter
3. Then perform inference based on experiment we see
   A: 1.3

C. Special Case: \( K = 2 \).
   1. \( X_1/X_+ \sim \mathcal{N}(\pi_1, \pi_1 \pi_0) \).
   2. \( p \)-value
      \( 2 \times \min (P [X_0 \geq x_0 | x_+ = X_+] , P [X_0 \leq x_0 | x_+ = X_+] ) \)
      a. If \( x_0 = 0 \), then \( P [X_0 \geq x_0] = 1 \): Never determines
         \( p \)-value
      b. If \( x_0 = x_+ \), then \( P [X_0 \leq x_0] = 1 \): Never determines
         \( p \)-value
      c. Using normal approximation,
      \( \approx 2 \min \left( \Phi \left( \frac{x_0 - \pi_0 x_+}{\pi_0 \pi_1 x_+} \right) , \Phi \left( \frac{-x_0 + \pi_0 x_+}{\pi_0 \pi_1 x_+} \right) \right) \)
      d. To properly account for probability at \( x_0 \), add \( \pm \frac{1}{2} \)
         to numerator to make absolute value smaller. See Figure 1.
      \( \approx 2 \min \left( \Phi \left( \frac{x_0 - \pi_0 x_+ + \frac{1}{2}}{\pi_0 \pi_1 x_+} \right) , \Phi \left( \frac{-x_0 + \pi_0 x_+ + \frac{1}{2}}{\pi_0 \pi_1 x_+} \right) \right) \) \( (1) \)

3. Get CI for \( \pi \) using
   a. Exact confidence bound
      i. Lower bound \( \pi_L \) satisfies \( P_{\pi} \geq x_0 = .025 \)
      ii. Upper bound \( \pi_U \) satisfies \( P_{\pi} \leq x_0 = .025 \)
      iii. Vertical line has probability .95 for any value of
         parameter
   b. Number of individuals in each category is given by a
   d. Poisson distribution with intensity \( \lambda_k \)
   i. \( P [X_k = x_k] = \exp(-\lambda_k)\lambda_k^{x_k} / x_k! \)
   e. Assume entries are independent
   f. Well known result: \( X_i \sim \mathcal{P}(\lambda_i) \), \( X_j \sim \mathcal{P}(\lambda_j) \) \( \Rightarrow \)
   \( X_i + X_j \sim \mathcal{P}(\lambda_i + \lambda_j) \).
   i. Obvious extension to more summands
   A: 1.2.1-1.2.2

Fig. 1: Demonstration of Continuity Correction

\textbf{Area approximated without cont. corr.}
\textbf{Additional area represented by cont. corr.}

\textbf{Potential Binomial Value}

iv. Hence horizontal line has same coverage
v. Lower confidence bound is generated by upper
   \textit{quantile} and \textit{vice versa}
vi. Figure 2 shows construction for a generic family
Lecture 1

vii. Figure 3 shows the construction for the binomial family. Note the role of discreteness.

Fig. 3: Graphical Construction of Confidence Interval For a Binomial Family

\[
\begin{align*}
\text{.025 quantiles} & \quad \text{Typical observed value} \quad \text{Confidence interval end points} \\
\hline
.025 & \quad 0.309 & \quad 0.4450 & \quad 0.5561 & \quad 0.653 & \quad 0.738 & \quad 0.813 \\
\hline
.975 & \quad 0.000 & \quad 0.0025 & \quad 0.0252 & \quad 0.067 & \quad 0.122 & \quad 0.187
\end{align*}
\]

Table 1: .95 Confidence Intervals for \(X_+ = 10\)

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal Lower</td>
<td>0.000</td>
<td>-0.086</td>
<td>-0.048</td>
<td>0.016</td>
<td>0.096</td>
<td>0.190</td>
</tr>
<tr>
<td>Normal Upper</td>
<td>0.000</td>
<td>0.2859</td>
<td>0.4479</td>
<td>0.584</td>
<td>0.704</td>
<td>0.810</td>
</tr>
<tr>
<td>Exact Lower</td>
<td>0.000</td>
<td>0.0025</td>
<td>0.0252</td>
<td>0.067</td>
<td>0.122</td>
<td>0.187</td>
</tr>
<tr>
<td>Exact Upper</td>
<td>0.309</td>
<td>0.4450</td>
<td>0.5561</td>
<td>0.653</td>
<td>0.738</td>
<td>0.813</td>
</tr>
</tbody>
</table>

b. Can be expressed in terms of \(F\) distribution upper quantile

\[
\left( X_0 + (X_1 + 1)F_{\alpha/2}(2X_0 + 2, 2X_1) \right) \quad X_1 + (X_0 + 1)F_{\alpha/2}(2X_0 + 2, 2X_1) \\
(Clopper and Pearson, 1934).
\]

i. Let \(G_{a,b}(y) = \int_0^y x^{a-1}(1-x)^b \, dx\) for continuous distributions in which quantile is data dependent.

\[
\hat{\theta} \approx N(\theta, \ell''(\theta)^{-1})
\]

i. Notation a bit abusive, since \(\ell''\) is data dependent.

5. Tests for some parameters and not others:

a. Notation:

i. \(\beta = (\psi, \phi)\)

ii. \(\psi\) of interest

iii. \(\phi\) not of interest

b. Likelihood Ratio:

i. Maximize \(\ell\) over \(\beta\)

ii. Maximize \(\ell\) with \(\psi = 0\) and \(\phi\) unconstrained.

iii. Compare \(2\times\) difference to \(\chi^2\)

- d.f. is length of \(\psi\)

c. Wald Test

i. Let \(\hat{\beta} = (\hat{\psi}, \hat{\phi})\) maximize \(\ell\) over \(\beta\)
ii. \( \text{Var} \left[ \hat{\beta} \right] \approx I(\hat{\beta}) = [-\ell''(\hat{\beta})]^{-1} \)

iii. \( \text{Var} \left[ \hat{\psi} \right] \approx \) appropriate submatrix \( I^{11} \)

iv. Wald statistic is \( \hat{\phi}^T I^{11}(\hat{\beta})^{-1} \hat{\phi} \)

d. Score Test
   i. Let \( \hat{\phi} \) maximize \( \ell \) over \( \phi \) with \( \psi = 0 \)
   ii. Let \( \ell' \) be components of \( \ell' \) corresponding to \( \psi \)
   iii. Test is \( \ell'(0, \bar{\psi})^T I^{11}(0, \bar{\psi}) \ell'(0, \bar{\psi}) \sim \chi^2 \)

6. Score test for binomial variables
   a. Score statistic is derivative of log of likelihood evaluated at null hypothesis
   b. \( L(\pi_0) = \left( \begin{array}{c} X_0 \\ X_1 \end{array} \right) \pi_0 X_0 (1 - \pi_0) X_1 \)
   c. \( l(\pi_0) = \log \left( \begin{array}{c} X_0 \\ X_1 \end{array} \right) \) + \log(\pi_0)X_0 + \log(1 - \pi_0)X_1
   d. \( l'(\pi_0) = X_0/\pi_0 - X_+ / (1 - \pi_0) = (1 - \pi_0)X_0 - \pi_0 (X_+ - X_0) = X_0 - \pi_0 X_+ / (1 - \pi_0) \)

7. Coverages shown in Figs. 4 and 5.

B&D2: 3.4c-e

D. Multiple (K) Categories
1. How do exposure groups differ?
2. Wrong Solution
   a. Choose one group as baseline
      i. Usually the one with no exposure, if there is one
      ii. Be careful what you lump in here
   b. Calculate relative risks with respect to this group
   c. Calculate hypothesis tests
      i. for each pair
      ii. or against a baseline

11 Lecture 2

\[
\ell(\lambda) = \frac{X_j}{\lambda_j} - \sum_{k=0}^{K-1} \frac{X_k}{\lambda_j} = (X_j - \lambda_j \sum_{k=0}^{K-1} \frac{X_k}{\lambda_j}) / \lambda_j = (X_j - \lambda_j X_+) / \lambda_j
\]

d. Second term is expected value
   i. Expectation in light of associated multinomial distribution
   ii. \( E_k = X_+ \lambda_k \)
   iii. When probabilities are proportional to something else, ex. time at risk: \( E_k = X_+ Q_k / \sum_j Q_j \)

e. Must reduce from K test statistics to one.

5. Unordered exposures
   a. Use as test statistic sum of score statistic components
      i. squared
      ii. weighted by estimated variance
      iii. \( \sum_k (X_k - E_k)^2 / E_k \)
   b. Distribution is that of sum of K squared \( \mathcal{N}(0,1) \)
      i. Not independent
      ii. Equivalent to \( K - 1 \) independent \( \mathcal{N}(0,1)^2 \)
      iii. Distribution called \( \chi^2 \) on \( K - 1 \) degrees of freedom; see Fig. 6.

\[
\sum_{k=0}^{K-1} X_k \log(\lambda_k) - \log(\lambda_+) + \log(C).
\]

c. Score statistics are
Fig. 6: Rejection Regions for Chi-Square Tests

\[ X_1 \]
\[ X_2 \]

\[ \alpha = 0.05, \text{ 3 category multinomial, 13 items} \]