I. Introduction

A: 1.1

A. Course investigates description and analysis of variables that are *categorical*

1. i.e., variables that name categories
2. Ex.,
   a. Healthy/Sick
   b. Young/Old
   c. Strongly Agree/Agree/Indifferent/Disagree/Strongly Disagree
   d. Alabama/Alaska/.../Wyoming

A: 1.1.2

3. Categories might be ordered or unordered
   a. Careful: can get fooled
      i. Ex., zip codes
   b. Key question will you (and how will you) use information about ordering

A: 1.1.1

4. May be used either as explanatory (independent) or response (dependent)
II. Inference on a Single Categorical Variable:
A. Distributional Assumptions:

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 $X_k$
   d. Number of individuals in each category is given by a 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.
Lecture 1

b. Conditional on observing \( X_+ = x_+ \),
\[
P \left[ X_0 = x_0, \ldots, X_{K-1} = x_{K-1} | X_+ = x_+ \right] = \frac{\prod_{k=0}^{K-1} \exp(-\lambda_k) \lambda_k^{x_k} / x_k!}{\exp(-\lambda_+) \lambda_+^{x_+} / x_+!} = \frac{K-1}{\prod_{k=0}^{K-1} \pi_k} \frac{x_+!}{x_0! \cdots x_{K-1}!}
\]
for \( \pi_k = \lambda_k / \lambda_+ \).

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: \( K = 2 \): \( X_0 | X_+ \sim \mathcal{N}(X_+ \pi_0, X_+ \pi_1 \pi_0) \).

B. Conditionality principal: If

1. data arises from random mixture of experiments
   a. Here indexed by \( X_+ \)

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. \( X_1 | X_+ \sim \mathcal{N}(X_+ \pi_1, X_+ \pi_1 \pi_0) \).

2. \( p \)-value
\[
2 \times \min(P \left[ X_0 \geq x_0 | x_+ = X_+ \right], P \left[ X_0 \leq x_0 | x_+ = X_+ \right])
\]
Lecture 1

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,

\[
\approx 2 \min \left( \Phi \left( \frac{x_0 - \pi_0 x_+}{\sqrt{\pi_0 \pi_1 x_+}} \right), \Phi \left( \frac{-x_0 + \pi_0 x_+}{\sqrt{\pi_0 \pi_1 x_+}} \right) \right)
\]

d. To properly account for probability at \( x_0 \), add \( \pm \frac{1}{2} \) to numerator to make absolute value smaller. See Figure 1.

\[
\approx 2 \min \left( \Phi \left( \frac{x_0 - \pi_0 x_+ + \frac{1}{2}}{\sqrt{\pi_0 \pi_1 x_+}} \right), \Phi \left( \frac{-x_0 + \pi_0 x_+ + \frac{1}{2}}{\sqrt{\pi_0 \pi_1 x_+}} \right) \right)
\]

3. Get CI for \( \pi \) using

a. Exact confidence bound

i. Lower bound \( \pi_L \) satisfies \( P_{\pi}[X_0 \geq x_0] = .025 \)

ii. Upper bound \( \pi_U \) satisfies \( P_{\pi}[X_0 \leq x_0] = .025 \)

iii. Vertical line has probability .95 for any value of parameter

iv. Hence horizontal line has same coverage

v. Lower confidence bound is generated by upper quantile and \textit{vice versa}

vi. Figure 2 shows construction for a generic family of continuous distributions in which quantile is non-decreasing in the parameter.
vii. Figure 3 shows the construction for the binomial family. Note the role of discreteness. [Mark b R]

[Mark c R]
b. Can be expressed in terms of $F$ distribution upper quantile
\[
\left( \frac{X_0}{X_0 + (X_1 + 1) F_{\alpha/2}(2X_1 + 2, 2X_0)}, \frac{(X_0 + 1) F_{\alpha/2}(2X_0 + 2, 2X_1)}{X_1 + (X_0 + 1) F_{\alpha/2}(2X_0 + 2, 2X_1)} \right)
\]
(Clopper and Pearson, 1934).

i. Let $G_{a,b}(y) = \int_0^y x^{a-1}(1 - x)^{b-1} \, dx \left( \frac{(a + b - 1)!}{(a - 1)!(b - 1)!} \right)$, for
Fig. 3: Graphical Construction of Confidence Interval For a Binomial Family

\[ a > -1. \]

ii. Suppose \( a > 0 \). Then
\[ G_{a+1,b}(y) = -y^a(1-y)^b \frac{(a+b)!}{a!b!} \]
\[ + \int_0^y x^{a-1}(1-x)^b \, dx \frac{(a+b)!}{(a-1)!b!} \]
\[ = -\binom{a+b}{b} y^a(1-y)^b + G_{a,b+1}(y). \]

iii. Then
\[ G_{n-b,b+1}(y) - G_{n-b+1,b}(y) = \binom{n}{b} y^{n-b}(1-y)^b. \]

iv. Also, \( G_{n,1}(y) = n \int_0^y x^{n-1} \, dx = y^n. \)

v. By induction, \( G_{n-b,b+1}(y) = \sum_{v=0}^b \binom{n}{v} y^{n-v}(1-y)^v. \)

vi. Finish using relationship between Beta and F distributions.

c. Normal approx. \( \pi \in X_1/X_+ \pm 1.96 \sqrt{\frac{X_0 X_1}{(X_1+X_0)^3}} \)

i. Problem if \( X_0 = 0 \)

ii. Less obvious problem for small \( X_1 + X_0 \)

d. Better Normal approx.: See homework.
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>

A: 3.5.2

4. Testing $H_0 : \beta = \beta^0$ vs $H_A : \beta \neq \beta^0$ using the likelihood function

a. $\hat{\beta} \approx \mathcal{N}(\beta, \ell''(\hat{\beta})^{-1})$

i. $\hat{\beta}$ approximately unbiased.

ii. $(\hat{\beta} - \beta)^\top [-\ell''(\hat{\beta})](\hat{\beta} - \beta)$ is *Wald test statistic*

b. $2 \times$ difference in max $\ell \approx \chi^2$

i. If $\ell$ maximized over two models, one nested in another,

ii. d.f. is difference in number of parameters

iii. if smaller model is correct.

iv. Value is *likelihood ratio test statistic*

c. $U(\beta) = \ell' \approx \mathcal{N}(0, -\ell'')$

i. Notation a bit abusive, since $\ell''$ is data dependent.
Lecture 1

d. $U^\top \left[ -\ell''(\beta) \right]^{-1}U$ is score test statistic

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. $\text{Var} \left[ \hat{\beta} \right] \approx I(\hat{\beta}) = \left[ -\ell''(\hat{\beta}) \right]^{-1}$
   iii. $\text{Var} \left[ \hat{\psi} \right] \approx$ appropriate submatrix $I^{11}$
   iv. Wald statistic is $\hat{\phi}^\top \left[ I^{11}(\hat{\beta}) \right]^{-1} \hat{\phi}$

d. Score Test
   i. Let $\tilde{\phi}$ maximize $\ell$ over $\phi$ with $\psi = 0$
   ii. Let $\ell^1$ be components of $\ell'$ corresponding to $\psi$
Lecture 1

iii. Test is \( l^1(0, \tilde{\eta})^\top I_{11}(0, \tilde{\eta}) l^1(0, \tilde{\eta}) \sim \chi^2 \)

6. Score test for binomial variables

a. Score statistic is derivative of log of likelihood evaluated at null hypothesis

b. \( L(\pi_0) = (\frac{X^+}{X_0})\pi_0^0X_0(1-\pi_0)X_1 \)

c. \( l(\pi_0) = \log(\frac{X^+}{X_0}) + \log(\pi_0)X_0 + \log(1-\pi_0)X_1 \)

d. \( l'(\pi_0) = \frac{X_0}{\pi_0} - X_+ - \frac{X_0}{(1-\pi_0)} = \frac{(1-\pi_0)X_0-\pi_0(X_+-X_0)}{(1-\pi_0)\pi_0} = \frac{X_0-\pi_0X_+}{(1-\pi_0)\pi_0} \)

7. Coverages shown in Figs. 4 and 5. [Mark e R]

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
Fig. 4: Coverage for One Sided 90% Confidence Interval

1
0.8
0.6
0.4
0.2
0
0 0.2 0.4 0.6 0.8 1
True probability

Sample size 10

i. for each pair

ii. or against a baseline

d. Claim heterogeneity if any of these shows up different

e. Problem of multiple comparisons

3. To avoid multiple comparisons, need one test for all groups
4. Choose a measure of disagreement with null answer

a. Use score test with multinomial model.

b. Like before, \( l(\lambda) = \sum_{k=0}^{K-1} X_k[\log(\lambda_k) - \log(\lambda_+)] + \log(C) \).

c. Score statistics are
\[ u_j(\lambda) = \frac{X_j}{\lambda} - \sum_{k=0}^{K-1} \frac{X_k}{\lambda} \]

\[ = \left( X_j - \lambda \sum_{k=0}^{K-1} \frac{X_k}{\lambda} \right) / \lambda \]

\[ = \left( X_j - \lambda \lambda_j X_+ \right) / \lambda_j \]

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.
Fig. 6: Rejection Regions for Chi-Square Tests

α = 0.05, 3 category multinomial, 13 items

6. [Mark g R] [Mark H R] [Mark H sas]