Inference for Two-Way Contingency Tables

Jinhwan Suk

Department of Mathematical Science, KAIST

June 16, 2020

Contents

Review

- Interval Estimation of Odds Ratio, Difference of Prop., RR
- Testing Independence (Multinomial sampling)
  1. Pearson χ^2 test (Under H_0 : independence)

$$X^2 = \sum_{i,j} \frac{(n_{ij} - \hat{\mu}_{ij})^2}{\hat{\mu}_{ij}} \xrightarrow{d} \chi^2_\nu$$

  2. Likelihood-ratio χ^2 test (Under H_0 : independence)

$$G^2 = -2 \log \Lambda = 2 \sum_{i,j} n_{ij} \log(n_{ij} / \hat{\mu}_{ij}) \xrightarrow{d} \chi^2_\nu$$

Review

- Testing Independence (Independent binomial sampling)
    1. Homogeneity condition $\pi_1 = \pi_2$

$$z = \frac{\hat{\pi}_1 - \hat{\pi}_2}{\sqrt{\hat{\pi}(1-\hat{\pi})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

- Score Confidence Intervals Comparing Proportions (Under $H_0 : \pi_1 - \pi_2 = \delta_0$)

$$z(\delta_0) = \frac{(\hat{\pi}_1 - \hat{\pi}_2) - \delta_0}{\left(\frac{\hat{\pi}_1(\delta_0)(1-\hat{\pi}_1(\delta_0))}{n_1} + \frac{\hat{\pi}_2(\delta_0)(1-\hat{\pi}_2(\delta_0))}{n_2}\right)^{1/2}}$$

Contents

Political Ideology	Happiness		
	Not too Happy	Pretty Happy	Very Happy
Liberal	13	29	15
Moderate	23	59	47
Conservative	14	67	54

- X and Y are ordinal. $\Longrightarrow \chi^2$ and G^2 chi-squared tests ignore !!
- How can we examine dependency??
  "Positive or Negative trend" $\leftarrow$ Correlation
- Assign **scores** $u_1 \leq u_2 \leq \cdots \leq u_I$ and $v_1 \leq v_2 \leq \cdots \leq v_J$

Two-Way Tables with Ordered Classifications

Linear Trend Alternative to Independence

Large sample normality of sample correlation coefficient.

$$
\begin{aligned}
\hat{\rho} &= \frac{\sum_1^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_1^n (X_i - \bar{X})^2 \sum_1^n (Y_i - \bar{Y})^2}} \\
&= \left(\frac{1}{n}\sum_1^n X_i^2 - \bar{X}^2\right)^{-1/2} \left(\frac{1}{n}\sum_1^n Y_i^2 - \bar{Y}^2\right)^{-1/2} \left(\frac{1}{n}\sum_1^n X_i Y_i - \overline{XY}\right) \\
&= (\bar{X^2} - \bar{X}^2)^{-1/2} (\bar{Y^2} - \bar{Y}^2)^{-1/2} (\overline{XY} - \bar{X}\bar{Y})
\end{aligned}
$$

Define $W_i = (X_i, Y_i, X_i^2, Y_i^2, X_i Y_i)$. Then, $\hat{\rho} = g(\bar{W})$ where

$$
g(u_1, u_2, u_3, u_4, u_5) = ((u_3 - u_1^2)(u_4 - u_2^2))^{-1/2}(u_5 - u_1 u_2).
$$

W.L.O.G. we may assume $\mathbb{E}X = \mathbb{E}Y = 0$ and $\mathbb{E}X^2 = \mathbb{E}Y^2 = 1$

By Central Limit Theorem, $\sqrt{n}(\bar{W} - (0, 0, 1, 1, \rho)) \xrightarrow{d} \mathcal{N}(0, cov(W_5))$

- H_0 : independent *vs.* $H_1 : \rho \neq 0$
- $\sqrt{n}\,\hat{\rho} \xrightarrow{d} \mathcal{N}(0,1)$ under H_0 (Delta Method)
- $M^2 = (n-1)\hat{\rho}^2$ is approximately χ^2_1
- Large p-value contradicts independence.
- Small p-value does not imply linear association.

Political Ideology	Happiness		
	Not too Happy	Pretty Happy	Very Happy
Liberal	13	29	15
Moderate	23	59	47
Conservative	14	67	54

- χ^2 test statistic : 7.07 **p-value : 0.13**
- Pearson Correlation(ρ) $= 0.135$
- M^2 statistic : 5.82 **p-value : 0.02**

- H_0 : Independent *vs.* H_1 : Monotone Trend
- (Method 1) : Assuming underlying continuous distribution
- (Method 2) : Using ordinal measure of association

$$\hat{\gamma} = \frac{\Pi_c}{\Pi_c + \Pi_d} - \frac{\Pi_d}{\Pi_c + \Pi_d}$$

- $\hat{\gamma}$ has approximately a normal sampling distribution.

$$z = \hat{\gamma} / SE$$

- Agresti (2010) : $(C - D)/SE_0$

- Cochran(1954) "any set of scores gives a valid test, provided that they are constructed **without consulting** the results of the experiment"
- The scale is chosen by a consensus of experts
- (1, 2, 3, 4) and (2, 4, 6, 8) have the same ρ and hence the same M^2

	Alcohol Consumption (average number of drinks per day)				
Malformation	0	<1	1-2	3-5	>5
Absent	17,066	14,464	788	126	37
Present	48	38	5	1	1

1. Score $=$ (0, 0.5, 1.5, 4.0, 7.0)

$$M^2 = 6.57 \quad \text{p-value : } 0.010$$

2. Score $=$ (1.0, 2.0, 3.0, 4.0, 5.0)

$$M^2 = 1.83 \quad \text{p-value : } 0.176$$

3. $Midrank = (8557.5, 24365, 32013, 32473, 32555.5)$

$$M^2(Spearman's\ rho) = 0.35 \quad \text{p-value : } 0.55$$

- Comparision of two groups in $2 \times J$ table
  $\Rightarrow$ *Wilcoxon* or *Mann-Whitney test*
- Linear trend test in $I \times 2$ table
  $\Rightarrow$ *Cochran-Armitage trend test* (Section 5.3.5)
- When one is norminal with more than 2 categories..?

Contents

- Sampling method : Poisson, Multinomial, Independent binomial
- **Assumption :** Individual observations are independent
- $(\{n_{i+}\}, \{n_{+j}\})$ is a sufficient statistic under H_0 : independent

  $\Rightarrow P(\{n_{ij}\}|\{n_{i+}\}, \{n_{+j}\})$ does not depend on unknown parameters
- In 2×2 table, the conditioning yields *hypergeometric distribution*
- For 2×2 table, independence is equivalent to $\theta = 1$

$$\text{P-value} = P(n_{11} \geq t_o) \cdots H_1 : \theta > 1$$

- The test for 2×2 tables is called *Fisher's exact test*.

	Guess Poured First		
Poured First	Milk	Tea	Total
Milk	3	1	4
Tea	1	3	4
Total	4	4	

- $H_0 : \theta = 1$ vs $H_1 : \theta > 1$
- P-value $= P(n_{11} \geq 3) = HG(8, 4, 3) + HG(8, 4, 4) = 0.243$
- As n_{11} larger,

  odds ratio, RR, and difference of proportions $\uparrow$

  $\Rightarrow$ SAME p-value

Small-Sample Inference for Contingency Tables

Two-Sided P-values for Fisher's Exact Test

1. (Irwin 1935) p-value $= \sum_{p(t) \leq p(t_0)} P(n_{11} = t)$
2. p-value $= \sum_{t \text{ is farther from } H_0 \text{ than } t_o} P(n_{11} = t)$

$$P\text{-value} = P[|n_{11} - \mathbb{E}n_{11}| \geq |t_0 - \mathbb{E}n_{11}|] = P(\chi^2 \geq \chi_o^2)$$

3. p-value $= 2\min[P(n_{11} \geq t_o), P(n_{11} \leq t_o)]$
4. (Blaker 2000) p-value $= \min[P(n_{11} \geq t_o), P(n_{11} \leq t_o)] + Q_0$

   Q_0 is the probability in other tail that is close to,
   but not greater than Q.

The approach of setting type 1 errors for one-sided tests at **half the conventional type 1 error** *used in two-sided tests is preferable in regulatory settings. This promotes consistency with two-sided confidence intervals that are generally appropriate for estimating the possible size of the difference between two treatments*

Small-Sample Inference for Contingency Tables

Discreteness and Conservatism Issues

- In previous example, possible one-sided p-values were restricted to

$$0.014, 0.243, 0.757, 0.986, 1.0$$

- 0.05 is not probable type 1 error $\Rightarrow$ Conservative
- Achieve any significance level by randomized test
- Tocher(1950) showed that Fisher's test is UMPU
- Or, Mid p-value $\Rightarrow$ Less Conservative

- Commonly, only $\{n_{i+}\}$ are fixed.
- Under binomial sampling, consider testing $H_0 : \pi_1 = \pi_2$

$$T = X^2$$

- Given $\pi_1 = \pi_2 = \pi$, the p-value is $P_\pi(T \geq t_0)$

$$P = \sup_{0 \leq \pi \leq 1} P_\pi(T \geq t_0)$$

- $T = X^2$ table : (3,0/0,3) p-value=?
- **Is it desirable to take supremum over all possible π?**

If the data X have a probability dist. P_ν and we wish to test $H_0 : \nu = \nu_0$, ...

Test statistic $T \Rightarrow$ p-value $= P(T \geq t_0)$

Now consider a model with a *nuisance parameter* θ, $X \sim P_{\nu,\theta}$...

Under $H_0 : \nu = \nu_0$, p-value $= \sup_\theta P_{\nu_0,\theta}(T \geq t_0)$

1. Choose T whose distribution under H_0 does not depend on θ

$$e.g. \quad \frac{\bar{X} - \mu_0}{\sigma} \Longrightarrow \frac{\bar{X} - \mu_0}{S}$$

   T is ancillary under H_0

2. Find a sufficient statistic S for θ under H_0 and condition on S.

$$p = P_{\nu_0}(T \geq t_0 | S = s) \quad \cdots \quad \text{Fisher's exact test}$$

   (Increases discreteness)

- Let C_γ be a $1-\gamma$ confidence set for θ under H_0.
- Restrict the maximization to C_γ

$$p_\gamma = \sup_{\theta \in C_\gamma} P_{\nu_0,\theta}(T \geq t_0) + \gamma$$

Definition (Valid p-value)

Valid p-value is a statistic p such that under the null hypothesis,

$$P(p \leq \alpha) \leq \alpha$$

Theorem

Suppose that $p(\theta)$ is a valid p-value for any assumed value θ. Let C_γ satisfy $P(\theta \in C_\gamma) \geq 1-\gamma$, if the null hypothesis is true. Then, p_γ is a valid p-value.

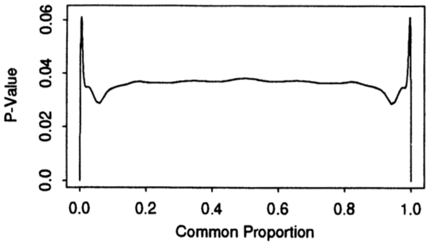
Proof :

$$\begin{aligned} P(p_\gamma \leq \alpha) &= P(p_\gamma \leq \alpha, \theta_0 \in C_\gamma) + P(p_\gamma \leq \alpha, \theta_0 \in \bar{C}_\gamma) \\ &\leq P(p(\theta_0) + \gamma \leq \alpha, \theta_0 \in C_\gamma) + P(\theta_0 \in \bar{C}_\gamma) \\ &\leq P(p(\theta_0) \leq \alpha - \gamma) + \gamma \\ &\leq \alpha - \gamma + \gamma \\ &= \alpha \end{aligned}$$

Example : Independence Test (Two independent binomial samples)

	Result		
Group	Success	Fail	Total
1	14	33	47
2	48	235	283

- $H_0 : \pi_1 = \pi_2$ and $T = Z^2 = \frac{(\hat{\pi}_1 - \hat{\pi}_2)^2}{\hat{\pi}(1-\hat{\pi})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$
- $p_{sup} = \sup_{\pi \in [0,1]} p(\pi) = .061$



- $\hat{\pi} = .188$ and $C_{.001} = [.123, .267]$

$$p_{.001} = .036 + .001 = .037$$

- $p(\hat{\pi})$ is typically not a valid p-value
- Maximization on the restricted set is much easier

Contents

Comparison of parameters for two independent binomial samples

$Y_i \sim bin(n_i, \pi_i), \;\; i = 1, 2, \; \pi_i \sim beta(\alpha_{i1}, \alpha_{i2})$ and π_1, π_2 independent
$\Rightarrow$ Independent posterior $beta(y_i + \alpha_{i1}, n_i - y_i + \alpha_{i2})$

- $H_0 : \pi_1 \leq \pi_2$ vs. $H_1 : \pi_1 > \pi_2$
  $\Rightarrow P(\pi_1 \leq \pi_2 | y_1, n_1; y_2, n_2)$: Bayesian p-value
- Intervals for difference of proportions, RR, Odds ratio

$$P(L < w < U) = F_w(U) - F_w(L)$$

$$F_w(t) = \int_{S_t} f(\pi_1, \pi_2 | y_1, n_1; y_2, n_2) \, d\pi_1 \, d\pi_2, \quad S_t = \{w \leq t\}$$

Bayesian Inference for Two-Way Contingency Tables

Example : Urn Sampling Gives Highly Unbalanced Treatment Allocation

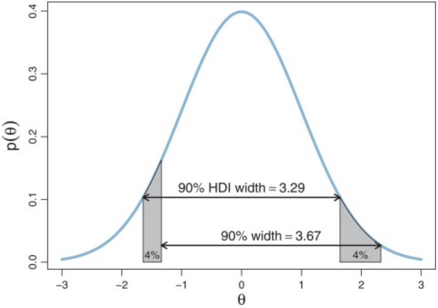
	Result		
Treatment	Success	Fail	Total
A	11	0	11
B	0	1	1

$$H_0 : \pi_1 = \pi_2 \ \ vs \ \ H_1 : \pi_1 > \pi_2$$

1. 95% *equal-tail* posterior interval for θ : (1.2, 218.4) (independent beta(2, 2) priors)
2. 95% *equal-tail* posterior interval for θ : (1.7, 4677) (independent uniform priors)
3. 95% *equal-tail* posterior interval for θ : $(3.3, 1.4 \times 10^6)$ (independent Jeffreys priors)
4. **Frequentist** : inverting large-sample score test $(4.5, \infty)$

Bayesian Inference for Two-Way Contingency Tables

Highest Posterior Density Intervals



- An alternative approach : Highest Posterior Density Interval(**HPDI**)
- Intuitively meaningful!! but, ...

Treatment	Result: Success	Result: Fail	Total
A	1	9	10
B	5	5	10

- HPDI for $\theta = (0.0006, 0.82)$
- HPDI for $1/\theta = (0.17, 38.23) \neq (1/0.82, 1/0.0006)$
- 95% ETI for $\theta = (0.017, 1.10)$
- 95% ETI for $1/\theta = (0.91, 57.9) = (1/1.10, 1/0.017)$

1. Bayesian Factor
   - Alternative to classical hypothesis testing(LRT)
   - Bayesian factor K is given by

$$K = \frac{P(D|H_0)}{P(D|H_1)}$$

2. Estimate Parameter
   - Estimate parameter that describe the association
   - correlation(ρ), Goodman-Kruskal's gamma(γ)

Contents

Exercise 3.3

	Homosexuals Should Have Right to Marry	
Political Party	Strongly Agree	Strongly Disagree
Strong Democrat	60	44
Strong Republican	2	61

- $\log(\hat{\theta}) = ?$
- Standard error of $\log(\hat{\theta}) = ?$
- The Wald 95% confidence interval for θ is ...?

$$\log(\hat{\theta}) = \log(60 * 61/2 * 44) = 3.728$$

$$\sigma^2 = \frac{1}{n_1} + \frac{1}{n_2} + \frac{1}{n_3} + \frac{1}{n_4} = 0.5558$$

$$(\exp(3.728 - 1.96 * 0.746), \exp(3.728 + 1.96 * 0.746)) = (9.6, 179.3)$$

n_{12} is too small! $\rightarrow$ interval estimate is so **imprecise**

Exercise 3.6

	Result		
Treatment	Success	Fail	Total
A	7	8	15
B	0	15	15
Total	7	23	

Q : Obtain a 95% confidence interval for the odds ratio

- The Wald CI :

$$\log(\hat{\theta}) = \infty, \quad \sigma^2 = \infty \Rightarrow \textit{Not exists}$$

- Profile Likelihood CI

$$G^2 = 2\sum_{i,j} n_{ij} \log(n_{ij}/\hat{\mu}_{ij}(\theta_0)) < \chi_1^2(.05)$$

Exercise 3.6 ▸ Code

- $\hat{\mu}_{ij}(\theta_0)$ is the unique expected frequency estimates that have the same row and column margins as $\{n_{ij}\}$ and satisfy

$$\frac{\hat{\mu}_{11}(\theta_0)\hat{\mu}_{22}(\theta_0)}{\hat{\mu}_{12}(\theta_0)\hat{\mu}_{21}(\theta_0)} = \theta_0$$

- Put $\hat{\mu}_{11}(\theta_0) = n_{\theta_0}$. Then,

$$\frac{n_{\theta_0}(n_{\theta_0}+8)}{(7-n_{\theta_0})(n_{\theta_0}+15)} = \theta_0$$

- We want collect all θ_0 such that

$$2\left(7\log\frac{7}{n_{\theta_0}} + 8\log\frac{8}{15-n_{\theta_0}} + 15\log\frac{15}{8+n_{\theta_0}}\right) < 3.841$$

- So we get $C_{.05} = (2.40, \infty)$