Topics in Statistics III/IV (MATH3361/4071) Michaelmas term, 2020

Handout 1: Introduction to 2 way contingency tables

Lecturer & author: Georgios P. Karagiannis georgios.karagiannis@durham.ac.uk

Aim

To get an introduction on how to read/write a contingency table and understand the sampling distributions

Notation: 2 way Contingency tables

Sampling distributions: Poisson, Product of Multinomials, Multinoimial

Reading list:

- Kateri, M. (2014; Chapters 1, 2). Contingency table analysis. Methods and implementation using R. Birkhauser
- Agresti, A. (2003; Chapters 1, 2, and 3). Categorical data analysis (Vol. 482). John Wiley & Sons.
- Lauritzen, S. L. (1996; Chapter 2). Graphical models (Vol. 17). Clarendon Press. [Maybe useful for Level 4]

1 $I \times J$ way contingency tables: notations, definitions

The set-up

- Let X and Y denote two categorical response variables (also called classifiers)
- X has I categories and Y with J categories.
- Classifications of subjects on both variables have IJ possible combinations.

The objective

- We are interested in performing inference on the probabilities
$$\pi_{ij} = \mathsf{P}(X = i, Y = j),$$
$i = 1, ..., I$, $j = 1, ..., J$ of responses (X, Y) of a subject chosen randomly from some population.
- E.g., the distribution $\pi_{ij} = \mathsf{P}(X = i, Y = j)$ of responses (X, Y) of a subject chosen randomly from some population.

We define the theoretical probabilities:

- $\pi_{ij} = \mathsf{P}(X = i, Y = j)$ is the probability getting an outcome $(X = i, Y = j)$
- $\pi_{i+} = \sum_j \pi_{ij} = \mathsf{P}(X = i)$ is the marginal probability of getting an outcome $X = i$ regardless the outcome of Y
- $\pi_{+j} = \sum_i \pi_{ij} = \mathsf{P}(Y = j)$ is the marginal probability of getting an outcome $Y = j$ regardless the outcome of X
- $\pi_{++} = \sum_i \sum_j \pi_{ij}$

One can tabulate the distribution $\pi_{ij} = \mathsf{P}(X = i, Y = j)$ of responses (X, Y) as in the following classification table (Table 1).

		Y					total
		1	$\cdots$	j	$\cdots$	J	
X	1	π_{11}	$\cdots$	π_{1j}	$\cdots$	π_{1J}	π_{1+}
	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
	i	π_{i1}	$\cdots$	π_{ij}	$\cdots$	π_{iJ}	π_{i+}
	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
	I	π_{I1}	$\cdots$	π_{Ij}	$\cdots$	π_{IJ}	π_{I+}
total		π_{+1}	$\cdots$	π_{+j}	$\cdots$	π_{+J}	π_{++}

Table 1: The distribution $\pi_{ij} = \mathsf{P}(X = i, Y = j)$ of responses (X, Y) in a IJ-table.

To learn the theoretical probabilities in Table 1, we collect a sample, (by using a sampling scheme from Section 2), and compute $(n_{i,j};\ \forall i = 1, ..., I,\ \forall j = 1, ..., J)$ where n_{ij} is the observed number of outcomes $(X = i, Y = j)$. We can specify related quantities, as:

- n_{ij} the observed number (or counts, frequency) of outcomes $(X = i, Y = j)$
- $n_{i+} = \sum_j n_{ij}$ is the observed marginal counts of outcomes $X = i$ regardless the outcome of Y
- $n_{+j} = \sum_i n_{ij}$ is the observed marginal counts of outcomes $Y = j$ regardless the outcome of X
- $n_{++} = \sum_{i,j} n_{ij}$ is the observed total counts (aka) the number of observations

Here, the $I \times J$ contingency table of X and Y [1] of observable frequencies $(n_{i,j})$, associated to the joint probabilities $(\pi_{i,j})$, is the rectangular table

- which has I rows for categories of X and J columns for categories of Y
- whose cells represent the IJ possible outcomes of the responses (X, Y)
- which displays $(n_{i,j})$ the observed number of outcomes $(X = i, Y = j)$ for each case

A schematic of the $(n_{i,j})$ contingency table displaying counts is presented in Table 2.

		Y					total
		1	$\cdots$	j	$\cdots$	J	
X	1	n_{11}	$\cdots$	n_{1j}	$\cdots$	n_{1J}	n_{1+}
	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
	i	n_{i1}	$\cdots$	n_{ij}	$\cdots$	n_{iJ}	n_{i+}
	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
	I	n_{I1}	$\cdots$	n_{Ij}	$\cdots$	n_{IJ}	n_{I+}
total		n_{+1}	$\cdots$	n_{+j}	$\cdots$	n_{+J}	n_{++}

Table 2: A $I \times J$ Contingency table of X and Y displaying the the observed number of outcomes $(X = i, Y = j)$

In a similar manner, a contingency table / classification table can display the following quantities

- $p_{i,j} = \frac{n_{i,j}}{n_{+,+}}$ the proportion of the outcomes $(X = i, Y = j)$,
- $\mu_{i,j}$ the expected number of outcomes $(X = i, Y = j)$,

[1] also called (n_{ij})-contingency table, or 2-way contingency table, or classification table

- $N_{i,j}$ the number of outcomes $(X = i, Y = j)$ (before the observations are collected)

When variable Y is treated as a response and variable X is treated as an explanatory variable, contingency tables can display the following conditional quantities

- $\pi_{i|j} = \frac{\pi_{ij}}{\pi_{+j}} = \mathsf{P}(X = i|Y = j)$: the conditional probability getting an outcome $X = i$ given that the outcome $Y = j$
- $\pi_{j|i} = \frac{\pi_{ij}}{\pi_{i+}} = \mathsf{P}(Y = j|X = i)$: the conditional probability getting an outcome $Y = j$ given that the outcome $X = i$
- $p_{j|i} = \frac{p_{ij}}{p_{i+}} = \frac{n_{ij}}{n_{i+}}$: the proportion of the outcomes $Y = j$, given that $X = i$
- $p_{i|j} = \frac{p_{ij}}{p_{+j}} = \frac{n_{ij}}{n_{+j}}$: the proportion of the outcomes $X = i$, given that $Y = j$

Two schematics of tables representing the conditional probabilities $\pi_{j|i}$, and conditional proportions $p_{j|i}$ of the outcomes $Y = j$, given that $X = i$ (aka conditioning on the rows) are in Tables 3 and 4.

$\pi_{j\mid i} = \frac{\pi_{ij}}{\pi_{i+}}$		Y					total
		1	$\cdots$	j	$\cdots$	J	
X	1	$\pi_{1\mid 1}$	$\cdots$	$\pi_{j\mid 1}$	$\cdots$	$\pi_{J\mid 1}$	$\pi_{+\mid 1} = 1$
	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
	i	$\pi_{1\mid i}$	$\cdots$	$\pi_{j\mid i} = \frac{\pi_{ij}}{\pi_{i+}}$	$\cdots$	$\pi_{J\mid i}$	$\pi_{+\mid i} = 1$
	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
	I	$\pi_{1\mid I}$	$\cdots$	$\pi_{j\mid I}$	$\cdots$	$\pi_{J\mid I}$	$\pi_{+\mid I} = 1$

Table 3: A table displaying the conditional proportions $\pi_{j|i}$ of the outcomes $Y = j$, given that $X = i$

$p_{j\mid i} = \frac{n_{ij}}{n_{i+}}$		Y					total
		1	$\cdots$	j	$\cdots$	J	
X	1	$p_{1\mid 1}$	$\cdots$	$p_{j\mid 1}$	$\cdots$	$p_{J\mid 1}$	$p_{+\mid 1} = 1$
	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
	i	$p_{1\mid i}$	$\cdots$	$p_{j\mid i} = \frac{n_{ij}}{n_{i+}}$	$\cdots$	$p_{J\mid i}$	$p_{+\mid i} = 1$
	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
	I	$p_{1\mid I}$	$\cdots$	$p_{j\mid I}$	$\cdots$	$p_{J\mid I}$	$p_{+\mid I} = 1$

Table 4: A table displaying the conditional proportions $p_{j|i}$ of the outcomes $Y = j$, given that $X = i$

Example 1. (Smoking vs. depression)

Consider a sample of size 3213 collected in 1980-1983 in St. Luis Epidemic Survey.
Assume Classification variables

X: Ever smoked (Yes, No)

Y: Major depression (Yes, No)

The observed counts are tabulated below.

	n_{ij}	Y		total
		yes	no	
X	yes	$n_{11} = 144$	$n_{12} = 1729$	$n_{1+} = 1873$
	no	$n_{21} = 50$	$n_{22} = 1290$	$n_{2+} = 1340$
total		$n_{+1} = 194$	$n_{+2} = 3019$	$n_{++} = 3213$

The joint proportions are

		Y		total
	p_{ij}	yes	no	
X	yes	$\frac{144}{3213}=0.04$	0.53	0.58
	no	0.01	0.4	0.41
total		0.06	0.93	1

The conditional proportions are

		Y		total
$p_{j\vert i}=\frac{n_{ij}}{n_{i+}}$		yes	no	
X	yes	$\frac{n_{11}}{n_{1+}}=\frac{144}{1873}=0.076$	0.92	1
	no	0.037	0.96	1

2 Sampling schemes

The observations (data set) are generated from sampling distributions specified by the sampling scheme (aka experiment) performed.

2.1 Poisson sampling scheme

Poisson sampling describes the scenario where the data-set (n_{ij}) is collected (sampled) independently in a given spatial, temporal, or other interval.

Then, n_{ij} is a realization of a Poisson random variable

$$N_{ij} \overset{\text{ind}}{\sim} \text{Poi}(\mu_{ij})$$

where

$$\mathsf{P}(N_{ij}=n_{ij}) = \frac{\exp(-\mu_{ij})\mu_{ij}^{n_{ij}}}{n_{ij}!}$$

with

$$\text{E}(N_{ij})=\mu_{ij} \qquad \text{and} \qquad \text{Var}(N_{ij})=\mu_{ij}$$

for $i=1,...,I$, and $j=1,...,J$.

The likelihood is

$$L(\boldsymbol{\mu}) \quad = \quad \prod_{\forall i,j}\frac{\exp(-\mu_{ij})\mu_{ij}^{n_{ij}}}{n_{ij}!} \quad \propto \quad \exp(-\sum_{\forall i,j}\mu_{ij})\prod_{\forall i,j}\mu_{ij}^{n_{ij}} \tag{1}$$

where $\boldsymbol{\mu}=(\mu_{11},\mu_{12},...,\mu_{IJ})$.

2.2 Multinomial sampling scheme

Multinomial sampling scheme describes the scenario where the data-set (n_{ij}) is collected (sampled) idependently given that the total sample size n_{++} is fixed (aka predetermined). Then $\boldsymbol{n}=(n_{11},n_{12},...,n_{IJ})$ is a realization of a Multinomial random variable <Appendix A

$$\underbrace{(N_{11},N_{12},...N_{IJ})}_{=\boldsymbol{N}} \overset{\text{ind}}{\sim} \text{Mult}(n_{++},\underbrace{(\pi_{11},...,\pi_{IJ})}_{=\boldsymbol{\pi}}) \tag{2}$$

So

$$\mathsf{P}(\boldsymbol{N}=\boldsymbol{n}|N_{++}=n_{++})=\frac{n_{++}!}{\prod_{\forall i,j} n_{ij}!}\prod_{\forall i,j}\pi_{ij}^{n_{ij}} \tag{3}$$

$\boldsymbol{N}=(N_{11},N_{12},...N_{IJ})$ and $\boldsymbol{\pi}=(\pi_{11},...,\pi_{IJ})$ with

$$\mathrm{E}(\boldsymbol{N})=n_{++}\boldsymbol{\pi} \qquad \text{and} \qquad \mathrm{Var}(\boldsymbol{N})=n_{++}(\mathrm{diag}(\boldsymbol{\pi})-\boldsymbol{\pi}\boldsymbol{\pi}^T)$$

The likelihood is

$$L(\boldsymbol{\pi}) \quad =\mathsf{P}(\boldsymbol{N}=\boldsymbol{n}|N_{++}=n_{++}) \quad =\frac{n_{++}!}{\prod_{\forall i,j} n_{ij}!}\prod_{\forall i,j}\pi_{ij}^{n_{ij}} \quad \propto \prod_{\forall i,j}\pi_{ij}^{n_{ij}} \tag{4}$$

as n_{++} is predetermined and fixed. Notice that if $\mu_{ij}=n_{++}\pi_{ij}$

$$L(\boldsymbol{\pi}) \quad \propto \prod_{\forall i,j}(\frac{\mu_{ij}}{n_{++}})^{n_{ij}} \quad \propto \prod_{\forall i,j}\mu_{ij}^{n_{ij}} \tag{5}$$

Relation between Multinomial and Poisson sampling.

We can produce the Multinomial distribution (describing the above experiment) based on the Poisson distr.

Remark 2. If r.v. $Y_i \sim \text{Poi}(\mu_i)$ for $i=1,...,k$ and they are independent then

$$\mathsf{P}(Y_1=y_1,...,Y_k=y_k|\sum_{i=1}^{k}Y_i=n)=\frac{n!}{\prod_{i=1}^{k}y_i!}\prod_{i=1}^{k}\pi_i^{y_i}1(y\in\mathcal{Y})$$

where $\mathcal{Y}=\left\{\boldsymbol{y}\in\{0,...,n\}^k|\sum_{j=1}^{k}Y_j=n\right\}$, and $\pi_i=\mu_i/n$.

Proof. It is easy to show that $\sum_{i=1}^{k}Y_i \sim \text{Poi}(\sum_{i=1}^{k}\mu_i)$ —(see Probability I). Then

$$\mathsf{P}(Y_1,...,Y_k|\sum_{i=1}^{k}Y_i=n)=\frac{\mathsf{P}(Y_1=y_1,...,Y_k=y_k)1(y\in\mathcal{Y})}{\mathsf{P}(\sum_{i=1}^{k}Y_i=n_{++})}=\frac{\prod_{\forall i}\frac{\exp(-\mu_i)\mu_i^{y_i}}{y_i!}1(y\in\mathcal{Y})}{\prod_{\forall i,j}\frac{\exp(-\sum_{i=1}^{k}\mu_i)(\sum_{i=1}^{k}\mu_i)^n}{n_{++}!}}$$

$$=\frac{n!}{\prod_{i=1}^{k}y_i!}\prod_{i=1}^{k}\underbrace{(\mu_i/n)}_{=\pi_i}{}^{y_i}1(y\in\mathcal{Y})$$

□

2.3 Product Multinomial sampling scheme

Product Multinomial sampling scheme describe the scenario where the sample (n_{ij}) is collected randomly and given that the marginal row size n_{i+} is fixed (aka predetermined) for $i=1,...,I$. Then $(n_{i1},n_{i2},...n_{iJ})$ is a realization of the Multinomial random variable

$$(N_{i1},N_{i2},...N_{iJ}) \overset{\text{ind}}{\sim} \text{Mult}(n_{i+},\boldsymbol{\pi}_i^*) \tag{6}$$

where $\boldsymbol{\pi}_i^*=(\pi_{1|i},...,\pi_{J|i})$ for $i=1,...,I$. In (6), 'ind' means that there is a Multinomial distribution for each row i and that rows are independent each other.

So

$$\mathsf{P}(\boldsymbol{N}=\boldsymbol{n}|\boldsymbol{N}_{i+}=\boldsymbol{n}_{i+}) = \prod_{\forall i}\left[\frac{n_{i+}!}{\prod_{\forall j} n_{ij}!}\prod_{\forall j}(\pi_i^*)^{n_{ij}}\right] \quad (7)$$

The likelihood is

$$L(\boldsymbol{\pi}) \quad = \prod_{\forall i}\left[\frac{n_{i+}!}{\prod_{\forall j} n_{ij}!}\prod_{\forall j}(\pi_i^*)^{n_{ij}}\right] \quad \propto \prod_{\forall i,j}\pi_{ij}^{n_{ij}}$$

as n_{i+} is predetermined and fixed. Notice that if $\mu_{ij} = n_{++}\pi_{ij}$

$$L(\boldsymbol{\pi}) \quad \propto \prod_{\forall i,j}(\frac{\mu_{ij}}{n_{++}})^{n_{ij}} \quad \propto (\frac{1}{n_{++}})^{\sum_{\forall i,j} n_{i,j}}\prod_{\forall i,j}\mu_{ij}^{n_{ij}} \quad \propto \prod_{\forall i,j}\mu_{ij}^{n_{ij}} \quad (8)$$

as the sum $\sum_{\forall i,j} n_{i,j} = \sum_{\forall i}\sum_{\forall i} n_{i,j} = \sum_{\forall j} n_{i,+}$ is fixed as a sum of predetermined marginal row sizes n_{i+}.

Remark 3. The three sampling schemes lead to the same MLE for $\boldsymbol{\mu}$'s or $\boldsymbol{\pi}$'s given that $\sum_{\forall i,j}\mu_{ij} = n_{++}$ is fixed. This is obvious in Multinomial sampling scheme and product Multinomial sampling scheme from 5 and 8. Regarding the Poisson, it is $\sum_{\forall i,j}\mu_{ij} = n_{++}$ and hence (1) becomes $L(\boldsymbol{\pi}) \propto \prod_{\forall i,j}\mu_{ij}^{n_{ij}}$.

2.4 Hyper-geometric sampling scheme

Hyper-geometric sampling scheme assumes the sample (n_{ij}) is collected randomly and given that the marginal counts n_{i+}, n_{+j} are fixed and predetermined before the experiment was performed.

The sampling distribution $\mathsf{P}(\boldsymbol{N}|N_{i+}=n_{i+}, N_{+j}=n_{+j})$ of $\boldsymbol{N}=(N_{11},...,N_{IJ})$ results by conditioning (3) properly and it is called Multivatiate Hyper geometric –we do not present the PMF here.

3 Types of independency in $I \times J$ tables

A number of important (in)dependencies modeled by $I \times J$ contingency tables:

1. X, Y are not independent iff

   $$\pi_{ij} = \text{well... have no particular structure...}, \quad \forall i, j$$

   So we need to learn the unknown $IJ-1$ quantities $\{\pi_{i,j}\}$

   This type of independency is symbolized as $[XY]$

2. X, Y are jointly independent iff

   $$\pi_{ij} = \pi_{i+}\pi_{+j}, \quad \forall i, j$$

   So to learn the unknown $\{\pi_{i,j}\}$, it is sufficient to learn just the $I+J-2$ unknown quantities $\{\pi_{i+}\}$, $\{\pi_{+j}\}$.

   This type of independency is symbolized as $[X, Y]$

4 Maximum likelihood estimation

Assume that the sampling distribution has been specified according to the sampling scheme used.

Assume that the relation of the classifiers X , Y has been specified/discovered.

Then the unknown probabilities $\{\pi_{i,j}\}$ can be learned via MLE.

To achieve this we Lagrange Multipliers can be used as a tool – (AMV II). <Appendix B

Example 4. Consider a Multinomial sampling scheme, and that X and Y are independent. Find the MLE of $\boldsymbol{\pi}$.

Solution. Because X and Y are independent, I have

$$\pi_{ij} = \pi_{i+}\pi_{+j}$$

The log likelihood is

$$\ell(\pi) = \sum_{ij} n_{ij} \log(\pi_{ij}) \; = \sum_{i} n_{i+} \log(\pi_{i+}) + \sum_{j} n_{+j} \log(\pi_{+j}) + \text{const}$$

Then Lagrange function is

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\pi}, \boldsymbol{\lambda}) &= \ell(\boldsymbol{\pi}) - \lambda_1(\sum_{i} \pi_{i+} - 1) - \lambda_2(\sum_{j} \pi_{+j} - 1) \\
&= \sum_{i} n_{i+} \log(\pi_{i+}) + \sum_{j} n_{+j} \log(\pi_{+j}) - \lambda_1(\sum_{i} \pi_{i+} - 1) - \lambda_2(\sum_{j} \pi_{+j} - 1)
\end{aligned}$$

It is

$$0 = \nabla_{\pi,\lambda} \mathcal{L}(\pi, \lambda)|_{(\pi,\lambda)=(\pi^*,\lambda^*)}$$

namely

$$\begin{cases}
0 = & \frac{\mathrm{d}}{\mathrm{d}\pi_{i+}} \mathcal{L}(\boldsymbol{\pi}, \boldsymbol{\lambda})|_{(\boldsymbol{\pi},\boldsymbol{\lambda})=(\boldsymbol{\pi}^*,\boldsymbol{\lambda}^*)} \\
0 = & \frac{\mathrm{d}}{\mathrm{d}\pi_{+j}} \mathcal{L}(\boldsymbol{\pi}, \boldsymbol{\lambda})|_{(\boldsymbol{\pi},\boldsymbol{\lambda})=(\boldsymbol{\pi}^*,\boldsymbol{\lambda}^*)} \\
0 = & \frac{\mathrm{d}}{\mathrm{d}\lambda_1} \mathcal{L}(\boldsymbol{\pi}, \boldsymbol{\lambda})|_{(\boldsymbol{\pi},\boldsymbol{\lambda})=(\boldsymbol{\pi}^*,\boldsymbol{\lambda}^*)} \\
0 = & \frac{\mathrm{d}}{\mathrm{d}\lambda_2} \mathcal{L}(\boldsymbol{\pi}, \boldsymbol{\lambda})|_{(\boldsymbol{\pi},\boldsymbol{\lambda})=(\boldsymbol{\pi}^*,\boldsymbol{\lambda}^*)}
\end{cases}
\overset{\text{calc.}}{\Longrightarrow}
\begin{cases}
\hat{\pi}_{i+} = \frac{n_{i+}}{n_{++}} \\
\hat{\pi}_{+j} = \frac{n_{+j}}{n_{++}} \\
\hat{\lambda}_1 = n_{++} \\
\hat{\lambda}_2 = n_{++}
\end{cases}$$

So

$$\hat{\pi}_{ij} = \hat{\pi}_{i+}\hat{\pi}_{+j} = \frac{n_{i+}}{n_{++}} \frac{n_{+j}}{n_{++}}$$

Example 5. Consider a Multinomial sampling scheme. We do not assume any independence for X and Y. We need to find the MLE of π.

The log likelihood is

$$\ell(\boldsymbol{\pi}) \propto \sum_{ij} n_{ij} \log(\pi_{ij})$$

Then Lagrange function is

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\pi}, \boldsymbol{\lambda}) &= \ell(\boldsymbol{\pi}) - \lambda(\sum_{ij} \pi_{ij} - 1) \\
&= \sum_{ij} n_{ij} \log(\pi_{ij}) - \lambda(\sum_{ij} \pi_{ij} - 1)
\end{aligned}$$

It is

$$0 = \nabla_{\boldsymbol{\pi},\boldsymbol{\lambda}} \mathcal{L}(\boldsymbol{\pi}, \boldsymbol{\lambda})|_{(\boldsymbol{\pi},\boldsymbol{\lambda})=(\boldsymbol{\pi}^*,\boldsymbol{\lambda}^*)}$$

namely

$$\dots \overset{\text{calc.}}{\Longrightarrow} \begin{cases} 0 = & \frac{\mathrm{d}}{\mathrm{d}\pi_{ij}}\mathcal{L}(\boldsymbol{\pi},\boldsymbol{\lambda})|_{(\boldsymbol{\pi},\boldsymbol{\lambda})=(\boldsymbol{\pi}^*,\boldsymbol{\lambda}^*)} \\ 0 = & \frac{\mathrm{d}}{\mathrm{d}\lambda_1}\mathcal{L}(\boldsymbol{\pi},\boldsymbol{\lambda})|_{(\boldsymbol{\pi},\boldsymbol{\lambda})=(\boldsymbol{\pi}^*,\boldsymbol{\lambda}^*)} \end{cases} \overset{\text{calc.}}{\Longrightarrow} \begin{cases} \hat{\pi}_{ij} = \frac{n_{ij}}{n_{++}} \\ \hat{\lambda} = n_{++} \end{cases}$$

So

$$\hat{\pi}_{ij} = \frac{n_{ij}}{n_{++}}$$

Example 6. [Cont. Example 1 (Smoking vs. depression)]
Compute the MLE of $\boldsymbol{\pi}$, given that X (Ever smoked) and Y (Major depression) are independent.

Solution. Based on the above discussion, it is

$$\hat{\pi}_{ij} = \hat{\pi}_{i+}\hat{\pi}_{+j} = \frac{n_{i+}}{n_{++}}\frac{n_{+j}}{n_{++}}$$

Hence

	$\hat{\pi}_{ij} = \frac{n_{i+}}{n_{++}}\frac{n_{+j}}{n_{++}}$	Y		total
		yes	no	
X	yes	0.03519801	0.5477463	0.5829443
	no	0.0251817	0.391874	0.4170557
total		0.06037971	0.9396203	1

5 Goodness of fit test (Model selection)

The pair of hypothesis test is as follows

$$\begin{cases} \mathrm{H}_0: & X, Y \text{ are independent} \\ \mathrm{H}_1: & X, Y \text{ are not independent} \end{cases} \Longleftrightarrow \begin{cases} \mathrm{H}_0: & \pi_{ij} = \pi_{i+}\pi_{+j} \\ \mathrm{H}_1: & X, Y \text{ are not independent} \end{cases}$$

The test can be based on the following two statistics:

- Pearson's statistic
$$X^2 \overset{\mathrm{H}_0}{=} \sum_{\forall i,j} \frac{(n_{ij} - \hat{\mu}_{ij})^2}{\hat{\mu}_{ij}} \underset{D}{\longrightarrow} \chi^2_{\text{df}}$$
- Likelihood ratio statistic
$$G^2 \overset{\mathrm{H}_0}{=} 2\sum_{\forall i,j} n_{ij}\log(\frac{n_{ij}}{\hat{\mu}_{ij}}) \underset{D}{\longrightarrow} \chi^2_{\text{df}}$$
- The degrees of freedom are
$$\text{df} = (I-1)(J-1)$$
and
$$\hat{\mu}_{ij} = \frac{n_{i+}n_{+j}}{n_{++}} : \text{the MLE of } \mu_{ij} \text{ under the model in } \mathrm{H}_0$$

The rejection areas at sig. level a are:

- For Pearson's statistic
$$R(\{n_{ij}\}) = \{X^2_{\text{obs}} \geq \chi^2_{\text{df},1-a}\}$$

- For Likelihood ratio statistic

$$R(\{n_{ij}\}) = \{G^2_{\text{obs}} \geq \chi^2_{\text{df},1-a}\}$$

The p-values are:

$$\text{pvalue} = \begin{cases} 1 - F_{\chi^2_{\text{df}}}(X^2_{\text{obs}}) & \text{, Pearson's statistic} \\ 1 - F_{\chi^2_{\text{df}}}(G^2_{\text{obs}}) & \text{, Likelihood ratio statistic} \end{cases}$$

Example 7. [Cont. Example 6(Smoking vs. depression)]
Perform hypothesis test to find out whether smoking a dispession are independent at sig. level 5%.

Solution. It is

$$\begin{cases} \text{H}_0: & X, Y \text{ are independent} \\ \text{H}_1: & X, Y \text{ are not independent} \end{cases} \Longrightarrow \begin{cases} \text{H}_0: & \pi_{ij} = \pi_{i+}\pi_{+j} \\ \text{H}_1: & X, Y \text{ are not independent} \end{cases}$$

- Based on Pearson test statistic I have: $X^2 = 21.557$, df $= 1$, $\chi^2_{1-0.05} = 3.841459$, and hence I reject the null hypothesis that smoking a dispersion are independent at sig. level. 5%
- Based on Pearson test statistic I have: $G^2 = 22.75$, df $= 1$, $\chi^2_{1-0.05} = 3.841459$, and hence I reject the null hypothesis that smoking a dispersion are independent at sig. level. 5%

6 Odds ratio in contingency tables

Before we go into the odds ratio in $I \times J$ tables recall concepts

Relative risk: Assume 2 events X and Y with probability of success π_X and π_Y respectively. The relative risk is

$$r = \frac{\pi_X}{\pi_Y}$$

Interpretation

$$\begin{aligned} r = 1 \Longrightarrow & \pi_X = \pi_Y \Longrightarrow & X \text{ is equaly likely to happen with } Y \\ r > 1 \Longrightarrow & \pi_X > \pi_Y \Longrightarrow & X \text{ is more likely to happen than } Y \\ r < 1 \Longrightarrow & \pi_X < \pi_Y \Longrightarrow & X \text{ is less likely to happen than } Y \end{aligned}$$

Odds: Assume 1 event X with probability of success π_X. The odds of X are

$$\Omega_X = \frac{\pi_X}{1 - \pi_X}$$

Interpretation

$$\begin{aligned} \Omega_X = 1 \Longrightarrow & \pi_X = 1 - \pi_X \Longrightarrow & \text{success is equally likely to happen} \\ \Omega_X > 1 \Longrightarrow & \pi_X > 1 - \pi_X \Longrightarrow & \text{success is more likely to happen} \\ \Omega_X < 1 \Longrightarrow & \pi_X < 1 - \pi_X \Longrightarrow & \text{success is less likely to happen} \end{aligned}$$

Odds ratio: Assume 2 events X and Y with probability of success π_X and π_Y respectively. The odds ratio of X vs Y are

$$\theta = \frac{\Omega_X}{\Omega_Y} = \frac{\pi_X/(1-\pi_X)}{\pi_Y/(1-\pi_Y)}$$

Interpretation

$$\begin{aligned}
\theta = 1 \Longrightarrow & \Omega_X = \Omega_Y \\
\theta > 1 \Longrightarrow & \Omega_X > \Omega_Y \\
\theta < 1 \Longrightarrow & \Omega_X < \Omega_Y
\end{aligned}$$

...more details on the interpretation will be given with respect to the contingency tables in what follows.

6.1 Odds ratio for 2×2 tables

Assume 2×2 tables

		Y		total
	$\pi_{i,j}$	1	2	
X	1	π_{11}	π_{12}	π_{1+}
	2	π_{21}	π_{22}	π_{2+}
total		π_{+1}	π_{+2}	π_{++}

$\Longrightarrow$

		Y		total
	$\pi_{j\vert i}$	1	2	
X	1	$\pi_{j=1\vert i=1}$	$\pi_{j=1\vert i=1}$	1
	2	$\pi_{j=1\vert i=2}$	$\pi_{j=2\vert i=2}$	1

The odds ratio of X vs Y are

$$\theta \quad = \frac{\Omega_X}{\Omega_Y} \quad = \frac{\pi_{j=1|i=1}(1-\pi_{j=2|i=1})}{\pi_{j=1|i=2}(1-\pi_{j=2|i=2})} \quad = \frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}}$$

Interpretation

- $\theta = 1 \iff \Omega_X = \Omega_Y \iff \pi_{j=1|i=1} = \pi_{j=2|i=1}$
  $\iff$ Y is equally likely to happen regardless whether $X = 1$ or $X = 2$
  $\iff$ X and Y are independent
- $\theta > 1 \iff \Omega_X > \Omega_Y \iff \pi_{j=1|i=1} > \pi_{j=2|i=1}$
  $\iff$ Y is more likely to happen if $X = 1$
  $\iff$ positive dependence
- $\theta < 1 \iff \Omega_X < \Omega_Y \iff \pi_{j=1|i=1} < \pi_{j=2|i=1}$
  $\iff$ Y is less likely to happen if $X = 1$
  $\iff$ negative dependence

The dependence becomes stronger as odds ratio θ moves away from 1

6.2 Odds ratio for $I \times J$ tables

- Decompose the $I \times J$ table into a minimal set of $(I-1)(J-1)$ 2×2 tables able to fully describe the problem in terms of odds ratio.
- However, this decomposition is not unique. Below, we present popular some choices.

Nominal Odds ratios

- Suitable for nominal classification variables
- They are defined in terms of a reference level, e.g. I-th and J-th
- The Nominal Odds ratios with reference levels I and J are

$$\theta_{ij}^{IJ} = \frac{\pi_{ij}/\pi_{iJ}}{\pi_{Ij}/\pi_{IJ}} = \frac{\pi_{ij}\pi_{IJ}}{\pi_{iJ}\pi_{Ij}}, \qquad \forall i = 1, ..., I-1; \qquad \forall j = 1, ...J-1$$

Local odds ratios

- Suitable for ordinal classification variables
- Compute each level of the ordinal classification to the immediate next
- The local Odds ratios are

$$\theta_{ij}^{L} = \frac{\pi_{i,j}/\pi_{i,j+1}}{\pi_{i+1,j}/\pi_{i+1,j+1}} = \frac{\pi_{i,j}\pi_{i+1,j+1}}{\pi_{i,j+1}\pi_{i,j+1}}, \qquad \forall i = 1, ..., I-1; \qquad \forall j = 1, ...J-1$$

$\Longrightarrow$

	j	$j+1$
i	$*$	$*$
$i+1$	$*$	$*$

- They can show us a taste of the trend
- They can recover all the odd rations, as:

$$\theta_{i,j}^{i+k,j+\ell} = \frac{\pi_{i,j}/\pi_{i,j+\ell}}{\pi_{i+k,j}/\pi_{i+k,j+\ell}} = \prod_{\rho=0}^{k-1}\prod_{\xi=0}^{\ell-1}\theta_{i+\rho,j+\xi}^{L} \qquad \forall k = 1, ..., I-i; \qquad \forall \ell = 1, ...J-j$$

$\Longrightarrow$

	j	$j+\ell$
i	$*$	$*$
$i+k$	$*$	$*$

- Note that $\theta_{ij}^{L} = \theta_{ij}^{i+1,j+1}$
- Note that for $k = I-i$ and $\ell = J-j$ we get the nominal odds ratios with reference levels I and J.

Global / cumulative odd ratios

- Suitable when I wish to treat the classification variables cumulatively.
- Global/cumulative odd ratios are a variation of the local odd ratios.
- They are the local odd ratios of collapsed/marginalised tables resulting by merging rows and/or columns of $I \times J$ tables.
- E.g.:

$$\begin{array}{|c|ccc|ccc|}
\hline
 & \multicolumn{6}{c|}{\text{Y}} \\
\hline
 & \pi_{11} & \cdots & \pi_{1j} & \cdots & & \pi_{1J} \\
 & \vdots & \ddots & \vdots & \ddots & & \vdots \\
\text{X} & \pi_{i1} & \cdots & \pi_{ij} & \cdots & & \pi_{iJ} \\
\cline{2-7}
 & \vdots & \ddots & \vdots & \ddots & & \vdots \\
 & \pi_{I1} & \cdots & \pi_{Ij} & \cdots & & \pi_{IJ} \\
\hline
\end{array}
\Longrightarrow
\begin{array}{|c|c|c|}
\hline
 & \multicolumn{2}{c|}{\text{Y}} \\
\hline
\text{X} & \sum_{l<i}\sum_{k\leq j}\pi_{lk} & \sum_{l<i}\sum_{k>j}\pi_{lk} \\
\cline{2-3}
 & \sum_{l>i}\sum_{k\leq j}\pi_{lk} & \sum_{l>i}\sum_{k>j}\pi_{lk} \\
\hline
\end{array}$$

$$\theta_{ij}^{\mathrm{G}} = \frac{\left[\sum_{l\leq i}\sum_{k\leq j}\pi_{lk}\right]\left[\sum_{l>i}\sum_{k>j}\pi_{lk}\right]}{\left[\sum_{l>i}\sum_{k\leq j}\pi_{lk}\right]\left[\sum_{l\leq i}\sum_{k>j}\pi_{lk}\right]},\ i=1,...,I-1,\ j=1,...,J-1$$

Maximum Likelihood Estimators

- The MLE estimate of $\theta_{i,j}^L$ is

$$\hat{\theta}_{i,j}^L = \frac{\hat{\pi}_{i,j}\hat{\pi}_{i+1,j+1}}{\hat{\pi}_{i,j+1}\hat{\pi}_{i+1,j}} = \frac{n_{i,j}, n_{i+1,j+1}}{n_{i,j+1}n_{i+1,j}}$$

- The asymptotic distribution of $\hat{\theta}_{i,j}^L$ (given that n_{++} is large) is such that

$$\frac{\log(\hat{\theta}_{i,j}^L) - \log(\theta_{i,j}^L)}{\sqrt{\frac{1}{n_{i,j}} + \frac{1}{n_{i,j+1}} + \frac{1}{n_{i+1,j}} + \frac{1}{n_{i+1,j+1}}}} \underset{D}{\longrightarrow} \mathrm{N}(0,1)$$

- The $(1-a)$ Confidence interval of $\log(\theta_{i,j}^L)$ is

$$(\log(\hat{\theta}_{i,j}^L) \pm z_{1-\frac{a}{2}}\sqrt{\frac{1}{n_{i,j}} + \frac{1}{n_{i,j+1}} + \frac{1}{n_{i+1,j}} + \frac{1}{n_{i+1,j+1}}})$$

- To test the hypothesis: $\mathrm{H}_0 : \theta_{i,j}^L = \theta_0$ v.s. $\mathrm{H}_0 : \theta_{i,j}^L \neq \theta_0$
  - rejection area: $\{\left|\frac{\log(\hat{\theta}_{i,j}^L)-\log(\theta_0)}{\sqrt{\frac{1}{n_{11}}+\frac{1}{n_{12}}+\frac{1}{n_{22}}+\frac{1}{n_{22}})}}\right| \geq z_{1-\frac{a}{2}}\}$
  - p-value= $2(1-P_{N(0,1)}(Z \leq \frac{\log(\hat{\theta}_{i,j}^L)-\log(\theta_0)}{\sqrt{\frac{1}{n_{i,j}}+\frac{1}{n_{i,j+1}}+\frac{1}{n_{i+1,j}}+\frac{1}{n_{i+1,j+1}})}}))$

  Likewise for other alternative hypothesis

Example 8. [Cont. Example 7(Smoking vs. depression)]
Compute the 95% confidence interval for the log odds ratio , and based on this test the hypothesis whether smoking a dispersion are independent at sig. level 5%.

Solution. Because I have just 2 levels for each categorical variable, the $(1-a)$ CI of $\log(\theta_{i,j}^L)$ is

$$(\log(\hat{\theta}) \pm z_{1-\frac{a}{2}}\sqrt{\frac{1}{n_{1,1}} + \frac{1}{n_{1,2}} + \frac{1}{n_{2,1}} + \frac{1}{n_{2,2}}})$$

So

$$\hat{\theta} = \frac{n_{11}, n_{2,2}}{n_{12}n_{2,1}} = 2.14...$$

$$\sqrt{\frac{1}{n_{1,1}} + \frac{1}{n_{1,2}} + \frac{1}{n_{2,1}} + \frac{1}{n_{2,2}}} = 0.16822$$

$$z_{1-\frac{0.05}{2}} = 1.96$$

Hence the 95% of $\log(\theta)$ is

$$(0.4351534\,,\ 1.094571)$$

To test the hypothesis whether smoking a dispersion are independent at sig. level 5% , I need to test

$$\mathrm{H}_0 : \theta = 1 v.s. \mathrm{H}_1 : \theta \neq 1$$

but

$$\left| \frac{\log(\hat{\theta}) - \log(1)}{\sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{22}} + \frac{1}{n_{22}})}} \right| = 4.54 > 1.96$$

and hence I reject the null hypothesis that smoking a dispersion are independent at sig. level 5%.

224

7 Analysis of residuals

If H_0 is rejected, we need to investigate which cell caused this rejection.
This can be done by analyzing the residuals and detecting large deviations.

- Pearson's residuals

$$e_{ij}^P = \frac{n_{ij} - \hat{\mu}_{ij}}{\sqrt{\hat{\mu}_{ij}}}$$

If Poisson sampling is performed then

$$e_{ij}^P \xrightarrow[D]{\mathrm{H}_0} \mathrm{N}(0, 1)$$

- Standardized residuals

$$\begin{aligned} e_{ij}^S &= e_{ij}^P \frac{1}{\sqrt{v_{ij}}} && (9) \\ &\approx e_{ij}^P \frac{1}{\sqrt{\hat{v}_{ij}}} = \frac{n_{ij} - \hat{\mu}_{ij}}{\sqrt{\hat{\mu}_{ij}}\sqrt{\hat{v}_{ij}}} && (10) \end{aligned}$$

where

$$\begin{aligned} v_{ij} &= (1 - \pi_{i+})(1 - \pi_{+j}) \\ \hat{v}_{ij} &= (1 - \frac{n_{i+}}{n_{++}})(1 - \frac{n_{+j}}{n_{++}}) \end{aligned}$$

Regardless the sampling scheme

$$e_{ij}^S \xrightarrow[D]{\mathrm{H}_0} \mathrm{N}(0, 1)$$

Why to standardize? The Pearson's residuals e_{ij}^S are not homogeneous if the sampling scheme is not Poisson, but $e_{ij}^P \xrightarrow[D]{\mathrm{H}_0} \mathrm{N}(0, v_{ij})$. Hence the adjustment in (9) standardizes them. Because v_{ij} are actually unknown parameters, they are approximated by their sampling analog in (10).

- Deviance residuals

$$e_{ij}^D = \mathrm{sign}(n_{ij} - \hat{\mu}_{ij}) \sqrt{2 n_{ij} \log(\frac{n_{ij}}{\hat{\mu}_{ij}})} \qquad (11)$$

where

$$e_{ij}^D \xrightarrow[D]{\mathrm{H}_0} \mathrm{N}(0, 1)$$

225

Criterion

If $|e_{ij}^S| > z_{1-\frac{a}{2}}$ then, at sig. level. a, the (i,j)th cell may be characterized as the influential causing the rejection of H_0.

- So you actually need to plot $\{e_{ij}^S\}$ against (i,j).
- Note that $|e_{ij}^S| > |e_{ij}^P|$ because ______________

Appendix

A Multinomial distribution

Assume random variables $\boldsymbol{Y} = (Y_1, ..., Y_k)^\top$ such that $Y_j \in \{0, ..., n\}$ and $\sum_{j=1}^{k} Y_j = n$. We say that $\boldsymbol{Y}$ follows a Multinomial distribution with parameters: number of trials $n > 0$ (integer), and success vector of probabilities $\boldsymbol{\pi} = (\pi_1, ..., \pi_k)^\top$ such that $\pi_i \in (0,1)$ and $\sum_{j=1}^{k} \pi_j = 1$ if:

$$\mathsf{P}\left(\{\boldsymbol{Y} = \boldsymbol{y}\}\right) = \frac{n!}{y_1!...y_k!}\pi_1^{y_1}...\pi_k^{y_k} 1\left(\boldsymbol{Y} \in \mathcal{Y}\right)$$

where $\mathcal{Y} = \left\{\boldsymbol{y} \in \{0, ..., n\}^k | \sum_{j=1}^{k} Y_j = n\right\}$.

- Notation

$$\boldsymbol{y} \sim \text{Mult}\left(n, \boldsymbol{\pi}\right)$$

- The Probability Mass Function is

$$f_{\boldsymbol{Y}}\left(\boldsymbol{y}\right) = \frac{n!}{y_1!...y_k!}\pi_1^{y_1}...\pi_k^{y_k} 1\left(Y \in \mathcal{Y}\right)$$

- The Expectation is

$$\mathsf{E}\left(\boldsymbol{Y}\right) = n\boldsymbol{\pi}$$
$$\mathsf{E}\left(Y_j\right) = n\pi_j, \quad \forall j$$

- The covariance matrix is such that

$$\text{Var}(\boldsymbol{Y}) = n\left(\text{diag}(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}^\top\right)$$
$$\text{Cov}(Y_j, Y_{j'}) = \begin{cases} n\pi_j(1-\pi_j), & j = j' \\ -n\pi_j\pi_{j'}, & j \neq j' \end{cases}, \quad \forall (j, j')$$

B Lagrange Multipliers (AMV II)

Problem: Let functions $f : \mathbb{R}^n \to \mathbb{R}$ and $g_k : \mathbb{R}^n \to \mathbb{R}$, for $k = 1, ..., K$.

$$\begin{cases} \text{maximize} & f(x) \\ \text{subject to} & g_k(x) = 0, \ \forall k = 1, ..., K \end{cases} \tag{12}$$

Assume ∇f, ∇g_1, ..., ∇g_K are continuous

Define Lagrange function is

$$\mathcal{L}(x, \lambda) = f(x) - \sum_{k=1}^{K} \lambda_k g_k(x)$$

where $\lambda = (\lambda_1, ..., \lambda_K)$.

Solution The solution x^* of (12) is the solution of

$$0 = \nabla_{x,\lambda}\mathcal{L}(x, \lambda)|_{(x,\lambda)=(x^*,\lambda^*)}$$