#### Math 362: Mathematical Statistics II

Le Chen

le.chen@emory.edu chenle02@gmail.com

> Emory University Atlanta, GA

Last updated on Spring 2021 Last compiled on January 15, 2023

2021 Spring

Creative Commons License (CC By-NC-SA)

# Chapter 10. Goodness-of-fit Tests

- § 10.1 Introduction
- § 10.2 The Multinomial Distribution
- § 10.3 Goodness-of-Fit Tests: All Parameters Known
- § 10.4 Goodness-of-Fit Tests: Parameters Unknown
- § 10.5 Contingency Tables

1

### Plan

- § 10.1 Introduction
- § 10.2 The Multinomial Distribution
- § 10.3 Goodness-of-Fit Tests: All Parameters Known
- § 10.4 Goodness-of-Fit Tests: Parameters Unknown
- § 10.5 Contingency Tables

## Chapter 10. Goodness-of-fit Tests

- § 10.1 Introduction
- § 10.2 The Multinomial Distribution
- § 10.3 Goodness-of-Fit Tests: All Parameters Known
- § 10.4 Goodness-of-Fit Tests: Parameters Unknown
- § 10.5 Contingency Tables

<i>p<sub>i</sub></i> are known	$p_i$ are unknown
$D = \sum_{i=1}^{t} \frac{(X_i - np_i)^2}{np_i}$	$D_1 = \sum_{i=1}^t rac{(X_i - n\hat{p}_i)^2}{n\hat{p}_i}$
$\chi^2$ with f.d. $\mathit{t}-1$	$\chi^2$ with f.d. $t-1-s$
$d = \sum_{i=1}^{t} rac{(k_i - np_{i0})^2}{np_{i0}}$	$d_1 = \sum_{i=1}^t rac{(k_i - n\hat{p}_{i0})^2}{n\hat{p}_{i0}}$
$np_{i0} \geq 5$	$\hat{np}_{i0} \geq 5$
$d>\chi^2_{1-lpha,t-1}$	$ extstyle d_1 > \chi^2_{1-lpha,t-1- extstyle s}$

† s is the number of unknown parameters.

 $\label{eq:df} \operatorname{\underline{df}} = \underline{\operatorname{number of classes}} - 1 - \operatorname{number of unknown parameters}.$ 

E.g. 1 Binomial data: 4096 students, each shots basketball 4 times. Let X<sub>i</sub> be the number of hits for the ith student.

People believe that  $X_i$  should following binomial (4, p), that is, shotting basketball should be something like trying to get red chocolate beans from a jar of beans of two colors.

Find the MLE for *p*. Use the data to make a conclusion.

- Sol. 1)  $H_0: X_i \sim \text{binomal}(4, p)$ .
  - 2) Under  $H_0$ , the MLE for p is  $p_e = ... = 0.251$

**E.g. 1** Binomial data: 4096 students, each shots basketball 4 times. Let  $X_i$  be the number of hits for the *i*th student.



Number of Hits, i	Obs. Freq., $k_i$
(0	1280
1	1717
$r_i's$ {2	915
3	167
4	17

Find the MLE for p. Use the data to make a conclusion.

Sol. 1)  $H_0: X_i \sim \text{binomal}(4, p)$ 

**E.g. 1** Binomial data: 4096 students, each shots basketball 4 times. Let  $X_i$  be the number of hits for the *i*th student.



Number of Hits, i	Obs. Freq., $k_i$
(0	1280
1	1717
$r_i's \{2$	915
3	167
4	17

Find the MLE for p. Use the data to make a conclusion.

Sol. 1)  $H_0: X_i \sim \text{binomal}(4, p)$ 

E.g. 1 Binomial data: 4096 students, each shots basketball 4 times. Let  $X_i$  be the number of hits for the *i*th student.



Number of Hits, i	Obs. Freq., $k_i$
(0	1280
1	1717
$r_i's \{2$	915
3	167
4	17

Find the MLE for p. Use the data to make a conclusion.

Sol. 1)  $H_0: X_i \sim \text{binomal}(4, p)$ .

**E.g. 1** Binomial data: 4096 students, each shots basketball 4 times. Let  $X_i$  be the number of hits for the *i*th student.



Number of Hits, i	Obs. Freq., $k_i$
(0	1280
1	1717
$r_i's$ 2	915
3	167
4	17

Find the MLE for p. Use the data to make a conclusion.

Sol. 1)  $H_0: X_i \sim \text{binomal}(4, p)$ .

$$\implies$$
  $d_1 = \cdots = 6.401.$ 

- 4) Critical region:  $(\chi^2_{95,5-1-1}, +\infty) = (7.815, +\infty)$
- 5) Conclusion: Fail to reject
- 6) Alternatively, *P*-value =  $\mathbb{P}(\chi_3^2 \ge 6.401) = 0.094$ , ... discuss...

Table 10.4.1		
Number of Hits, i	Obs. Freq., $k_i$	Estimated Exp. Freq., $n \hat{p}_{i_o}$
$r_i's$ $\begin{cases} 0\\1\\2\\3\\4 \end{cases}$	1280 1717 915 167	1289.1 1728.0 868.6 194.0 16.3

$$\implies$$
  $d_1 = \cdots = 6.401.$ 

- 4) Critical region:  $(\chi^2_{.95,5-1-1}, +\infty) = (7.815, +\infty)$
- 5) Conclusion: Fail to reject
- 6) Alternatively, *P*-value =  $\mathbb{P}(\chi_3^2 \ge 6.401) = 0.094$ , ... discuss...

Table 10.4.1		
Number of Hits, i	Obs. Freq., $k_i$	Estimated Exp. Freq., $n \hat{p}_{i_o}$
$r_i's$ $\begin{cases} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{cases}$	1280 1717 915 167 17	1289.1 1728.0 868.6 194.0 16.3

$$\implies$$
  $d_1 = \cdots = 6.401.$ 

- 4) Critical region:  $(\chi^2_{.95,5-1-1}, +\infty) = (7.815, +\infty)$
- 5) Conclusion: Fail to reject.
- 6) Alternatively, *P*-value =  $\mathbb{P}(\chi_3^2 \ge 6.401) = 0.094$ , ... discuss...

Table 10.4.1		
Number of Hits, i	Obs. Freq., $k_i$	Estimated Exp. Freq., $n \hat{p}_{i_o}$
$r_i's$ $\begin{cases} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{cases}$	1280 1717 915 167 17	1289.1 1728.0 868.6 194.0 16.3

$$\implies$$
  $d_1 = \cdots = 6.401.$ 

- 4) Critical region:  $(\chi^2_{.95,5-1-1}, +\infty) = (7.815, +\infty)$
- 5) Conclusion: Fail to reject.
- 6) Alternatively, *P*-value =  $\mathbb{P}(\chi_3^2 \ge 6.401) = 0.094$ , ... discuss...

E.g. 2 Does the number of death per day follow the Poisson distribution?

E.g. 2 Does the number of death per day follow the Poisson distribution?

Number of Deaths, i	Obs. Freq., $k_i$
0	162
1	267
2	271
2 3	185
4	111
5	61
6	27
7	8
8	3
9	1
10+	0
	1096

- 2)  $H_0: X_i$  follow Poisson( $\lambda$ ).
- 3) The MLE for  $\lambda$  is:  $\lambda_e = \cdots = 2.157$
- 4) Compute the expected frequencies:

- Sol. 1) Let  $X_i$  be the number of death in ith day,  $1 \le i \le 1096$ .
  - 2)  $H_0: X_i$  follow Poisson( $\lambda$ ).
  - 3) The MLE for  $\lambda$  is:  $\lambda_e = \cdots = 2.157$ .
  - 4) Compute the expected frequencies:

- 2)  $H_0: X_i$  follow Poisson( $\lambda$ ).
- 3) The MLE for  $\lambda$  is:  $\lambda_e = \cdots = 2.157$ .
- 4) Compute the expected frequencies:

- Sol. 1) Let  $X_i$  be the number of death in ith day,  $1 \le i \le 1096$ .
  - 2)  $H_0: X_i$  follow Poisson( $\lambda$ ).
  - 3) The MLE for  $\lambda$  is:  $\lambda_e = \cdots = 2.157$ .
  - 4) Compute the expected frequencies:

$$\implies$$
  $d_1 = \cdots = 25.98.$ 

2)  $H_0: X_i$  follow Poisson( $\lambda$ ).

3) The MLE for  $\lambda$  is:  $\lambda_e = \cdots = 2.157$ .

Table 10.4.2		
Number of Deaths, i	Obs. Freq., $k_i$	Est. Exp. Freq., $n  \hat{p}_{i_o}$
0	162	126.8
1	267	273.5
2	271	294.9
3	185	212.1
4	111	114.3
5	61	49.3
6	27	17.8
7		5.5
8		1.4
9		0.3
10+		0.1
	1096	1096

$ \begin{matrix} 0 & 162 & 126.8 \\ 1 & 267 & 273.5 \\ 2 & 271 & 294.9 \\ r_1, r_2, \dots, r_8 & 3 & 185 & 212.1 \\ 4 & 111 & 114.3 \\ 5 & 61 & 49.3 \\ \end{matrix} $			Table 10.4.3
$ \begin{vmatrix} 1 & 267 & 273.5 \\ 2 & 271 & 294.9 \\ 3 & 185 & 212.1 \\ 4 & 111 & 114.3 \\ 5 & 61 & 49.3 \\ \end{vmatrix} $	Exp. Freq., $n \hat{p}_{i_o}$	s, $i$ Obs. Freq., $k_i$	Number of Deaths, i
	114.3 49.3 17.8 7.3	267 271 185 111 61 27	$r_1, r_2, \dots, r_8$ $\begin{cases} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{cases}$

$$\implies$$
  $d_1 = \cdots = 25.98.$ 

2)  $H_0: X_i$  follow Poisson( $\lambda$ ).

3) The MLE for  $\lambda$  is:  $\lambda_e = \cdots = 2.157$ .

Table 10.4.2		
Number of Deaths, i	Obs. Freq., $k_i$	Est. Exp. Freq., $n  \hat{p}_{i_o}$
0	162	126.8
1	267	273.5
2	271	294.9
3	185	212.1
4	111	114.3
5	61	49.3
6	27	17.8
7		5.5
8		1.4
9		0.3
10+		0.1
	1096	1096

Table 10.4.3						
Number of Deaths, i		Obs. Freq., k <sub>i</sub>	Est. Exp. Freq., $n\hat{p}_{i_0}$			
$r_1, r_2, \ldots, r_8$	0 1 2 3 4 5 6 7+	162 267 271 185 111 61 27 12	126.8 273.5 294.9 212.1 114.3 49.3 17.8 7.3			
		1096	1096			

$$\implies$$
  $d_1 = \cdots = 25.98.$ 

- Sol. 1) Let  $X_i$  be the number of death in ith day,  $1 \le i \le 1096$ .
  - 2)  $H_0: X_i$  follow Poisson( $\lambda$ ).
  - 3) The MLE for  $\lambda$  is:  $\lambda_e = \cdots = 2.157$ .
  - 4) Compute the expected frequencies:

Table 10.4.2				
Number of Deaths, i	Obs. Freq., k <sub>i</sub>	Est. Exp. Freq., $n \hat{p}_{i_o}$		
	162	126.8		
	267	273.5		
	271	294.9		
	185	212.1		
	111	114.3		
	61	49.3		
	27	17.8		
		5.5		
		1.4		
		0.3		
10+		0.1		
	1096	1096		

Table 10.4.3					
Number of Deaths, i		Obs. Freq., k <sub>i</sub>	Est. Exp. Freq., $n\hat{p}_{i_0}$		
$r_1, r_2, \ldots, r_8$	0 1 2 3 4 5 6 7+	162 267 271 185 111 61 27 12	126.8 273.5 294.9 212.1 114.3 49.3 17.8 7.3		
		1096	1096		

$$\implies$$
  $d_1 = \cdots = 25.98.$ 

5) *P*-value = 
$$\mathbb{P}(\chi_{1.8-1-1}^2 \ge 25.98) = 0.00022$$
. Reject!