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

Chapter 10. Goodness-of-fit Tests

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

† s is the number of unknown parameters.

$$\text{df} = \underline{\text{number of classes}} - 1 - \underline{\text{number of unknown parameters.}}$$

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

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

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) $H_0 : X_i \sim$binomal$(4, p)$.

2) Under H_0, the MLE for p is $p_e = ... = 0.251$

3) Compute the expected frequenies:

Table 10.4.1

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

$$\Longrightarrow \quad d_1 = \cdots = 6.401.$$

4) Critical region: $(\chi^2_{.95,5-1-1}, +\infty) = (7.815, +\infty)$

5) Conclusion: Fail to reject.

6) Alternatively, P-value $= \mathbb{P}(\chi^2_3 \geq 6.401) = 0.094$, ... discuss... $\square$

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
3	185
4	111
5	61
6	27
7	8
8	3
9	1
10+	0
	1096

Sol. 1) Let X_i be the number of death in ith day, $1 \leq i \leq 1096$.

2) H_0 : X_i follow Poisson(λ).

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

4) Compute the expected frequencies:

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	8	5.5
8	3	1.4
9	1	0.3
10+	0	0.1
	1096	1096

Table 10.4.3

	Number of Deaths, i	Obs. Freq., k_i	Est. Exp. Freq., $n\hat{p}_{i_o}$
$r_1, r_2, \ldots, r_8$	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+	12	7.3
		1096	1096

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

5) P-value $= \mathbb{P}(\chi^2_{1,8-1-1} \geq 25.98) = 0.00022$. Reject! $\square$