

Data Analysis (046193) HW3

Submission date: 25/5/17

- * Submission in pairs only!
- * Submit a ZIP file containing your files named with 9 digits of your IDs. Submission Example: "200567989_123456789.zip".
- * Submit your computer exercise solution as an "ipython notebook", after you zip your files extract them into a new folder and make sure there are no runtime errors.
- * Use python version 2.7.

Hypothesis testing

1. Complete the proof from the lecture notes page 2-27
2. Show that the Pearson Chi-square test and the likelihood ratio Chi-squared test converge to the same test as N approaches infinity.
 - a. Use Wilks theorem and show that the likelihood ratio converges to a Chi-Square distribution.
 - b. We will now prove that the Pearson test also converges to the same Chi-square distribution.
 - i. Use the CLT theorem and show that the random variable holds the following:

$$z_i = \frac{X_i - np_i}{\sqrt{np_i}} \rightarrow N(0, 1 - p_i)$$

Where X,n,p are as they defined in class (page 2-33).

- ii. Show that $Ez_i z_j = -\sqrt{p_j p_i}$
3.
 - a. A 6-sided die is thrown 60 times. The number of times it lands with 1, 2, 3, 4, 5 and 6 face up is 8,4,11,9,11 and 17, respectively. Is the die biased, according to the Pearson's chi-square test and the log-likelihood chi-square test at a significance level $(1-\alpha)$ of:
 - i. 99%?
 - ii. 95%?
 - iii. What is the p-value for this test?

- b. Now the die is thrown 600 times and the results are 80,40,110,90,110 and 170. Repeat the experiment. What can you say on the results with respect to the previous question?

Hint- use a Chi-Square table for the relevant number of degrees of freedom.

4. Let $X_1, \dots, X_n \sim \text{Poisson}(\lambda)$

- a. Let $\lambda_0 > 0$. Find the size α Wald test for:

$$H_0: \lambda = \lambda_0 \text{ Versus } H_1: \lambda \neq \lambda_0$$

- b. **Computer.**

Let $\lambda_0 = 1$, $n = 20$, $\alpha = 0.05$

Simulate $X_1, \dots, X_n \sim \text{Poisson}(\lambda_0)$

Perform the Wald test

Repeat many times and count how often you reject the null.

How close is the type 1 error rate to α ?

5.

שאלה 5- ממבחן מועד ב' אביב תשע"ו:

ב- 1000 הטלות בלתי-תלויות של קובייה בת שלוש פאות התקבלו התוצאות הבאות (N_j) הוא מספר ההטלות שבהן התקבלה פאה j :
 $N_1 = 350, N_2 = 340, N_3 = 310$
 נדרש לבחון באם הקובייה היא הוגנת.

א. נסחו בעיית בדיקת השערות מתאימה. יש להגדיר באופן מלא את הפילוגים והפרמטרים המתאימים.

ב. רישמו את פונקציית הסבירות $L(\theta)$, ואת יחס הסבירות $\frac{L(\theta)}{L(\theta_0)}$.

ג. יישמו את מבחן יחס-הסבירות למקרה זה. יש לקבל ביטוי מפורש ככל האפשר (כתלות בקבוע c).

ד. הציעו מבחן מתאים נוסף לבעייה זו.

ה. נדרש כי המבחן יהיה בעל גודל $\alpha = 0.05$. הסבירו כיצד תבחרו את הסף c כך שהמבחן של סעיף ג או ד (לבחירתכם) יקיים זאת בקירוב.

הנחייה: הניחו כי לרשותכם טבלאות של הפילוג המצרפי F_d של משתנים בבעלי פילוג χ_d^2 לפי הצורך. אין צורך בתשובה מספרית, אך יש לפרט באופן מלא את אופן החישוב.

ו. חשבו את ערך- κ של המבחן שהצעתם בסעיף ג או ד.
הנחיה: ראו הסעיף הקודם.

Dimensionality reduction

6. Computer.

- a. Load the connectionist bench data set from
[https://archive.ics.uci.edu/ml/datasets/Connectionist+Bench+\(Sonar,+Mines+vs.+Rocks\)](https://archive.ics.uci.edu/ml/datasets/Connectionist+Bench+(Sonar,+Mines+vs.+Rocks))
- b. Give a short explanation on the data set, you may include some visualization of it (similar to the tutorial).
- c. Compute its PCA representation.
- d. Visualize the data on 2d and 3d scatter plots.
- e. Present the error curve.
- f. How many Eigen vectors can you take to obtain:
 - i. 95% error?
 - ii. 99%?