

# Problem Set 2

*Yingxin Cao*

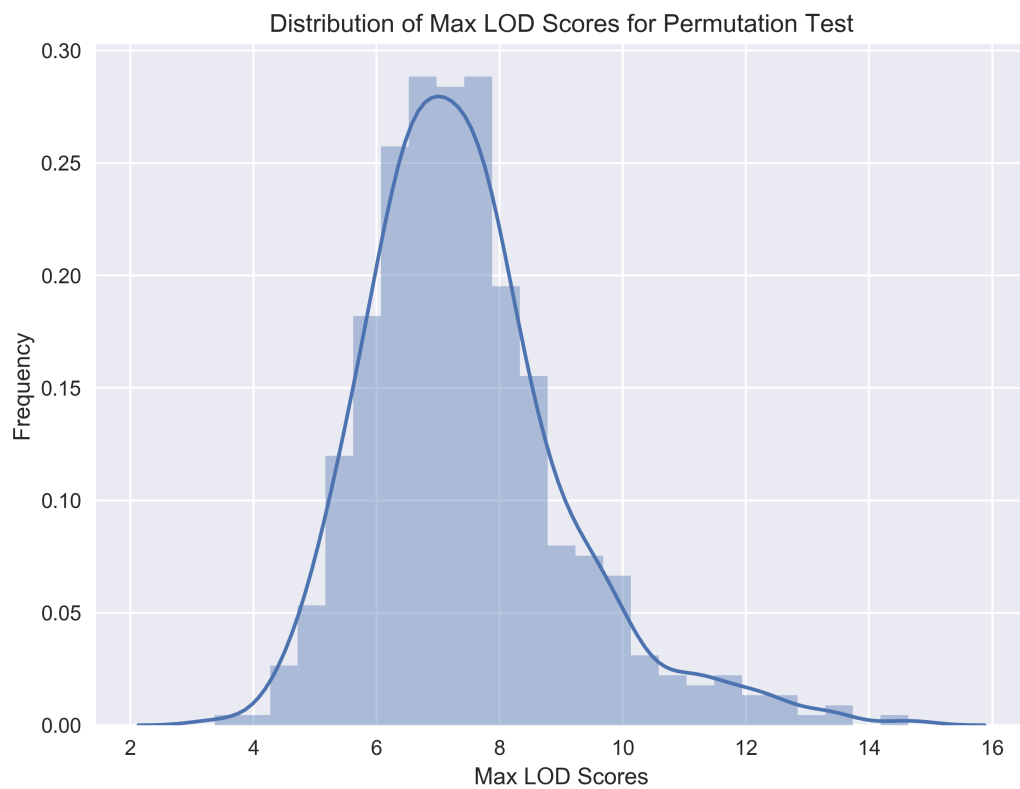
## Problem 1.

a)

The max LOD score is 18.2623422045. Please see .py file for implementation.

b)

Set the permutation size to be 500. The result is shown below.



The 95% threshold is 10.5652802439

c)

The NO. of genetic markers that have a score greater than threshold are

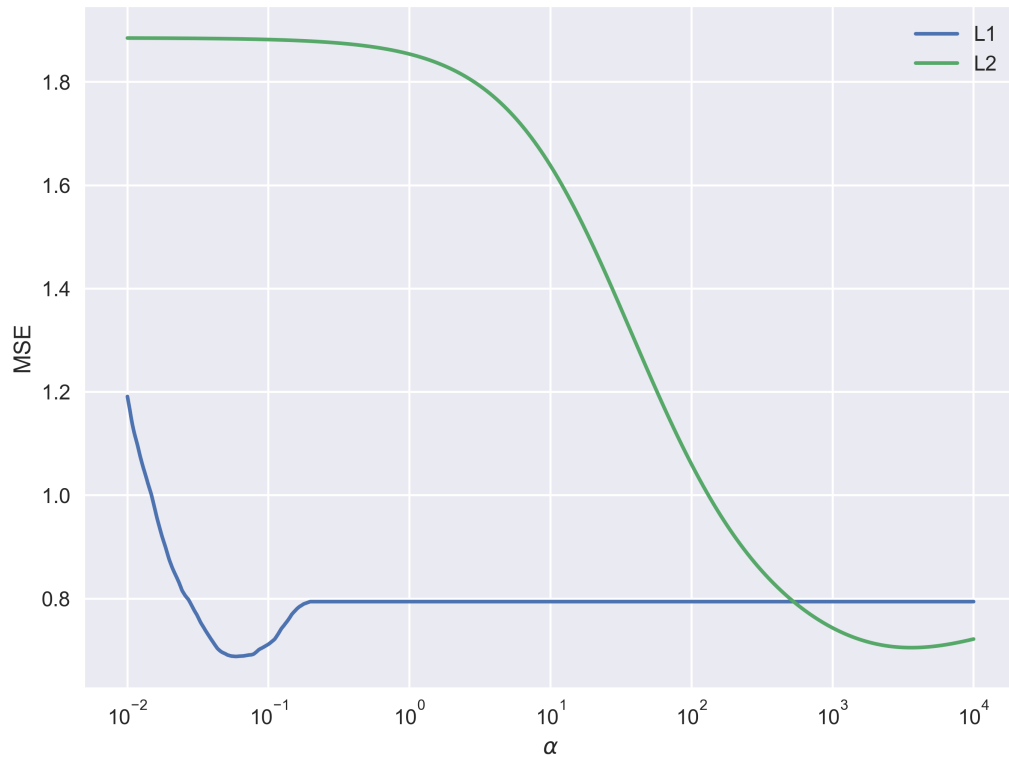
[ 43 124 273 561 754 846 1123]

## Problem 2.

a)

The range is defined by

```
alplt_range = np.logspace(-2, 4, 1000)
```



b)

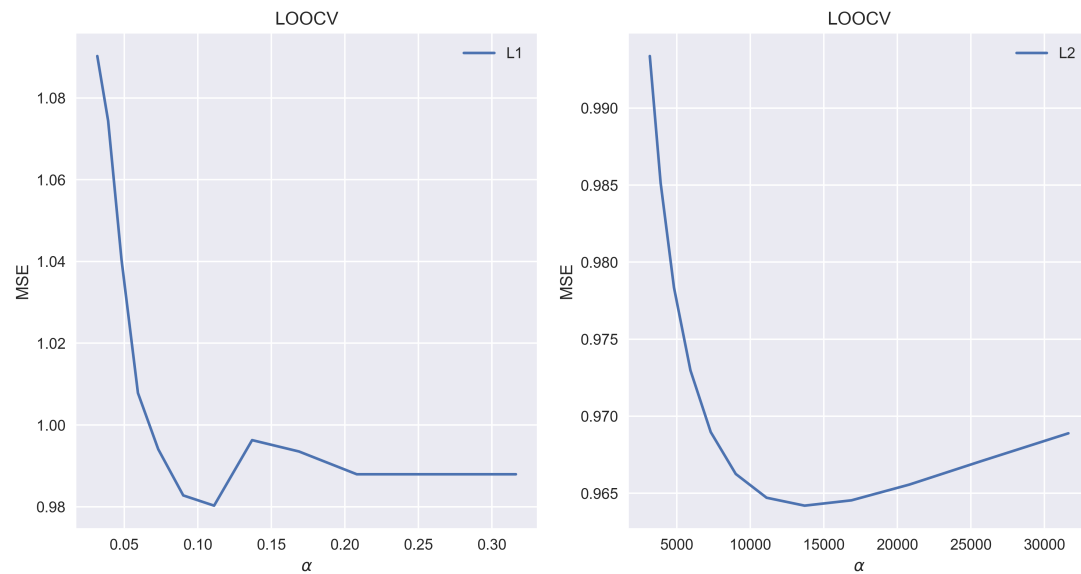
The result is also shown in the above figure.

Min MSE for lasso is 0.688103854939. Min MSE for ridge is 0.70514728884.

L2 has higher MSE than L1.

c)

Due to the computation cost, less points are selected.



min MSE for lasso is 0.980266768121 min MSE for ridge is 0.96419177019

According to the result. L2 does better.

The MSE are higher using LOOCV. The difference is around 0.27.

### Problem 3.

a)

Hidden variables: haplotypes;

Parameters: frequencies.

b)

$$z_{11} = \frac{1/12 \times 2/12}{1/12 \times 2/12 + 3/12 \times 1/12} = 0.4 \quad z_{12} = \frac{3/12 \times 1/12}{1/12 \times 2/12 + 3/12 \times 1/12} = 0.6$$

$$z_{21} = \frac{1/12 \times 3/12}{1/12 \times 3/12 + 1/12 \times 1/12} = 0.75 \quad z_{22} = \frac{1/12 \times 1/12}{1/12 \times 3/12 + 1/12 \times 1/12} = 0.25$$

$$z_{31} = \frac{3/12 \times 1/12}{3/12 \times 1/12 + 2/12 \times 1/12} = 0.6 \quad z_{32} = \frac{2/12 \times 1/12}{3/12 \times 1/12 + 2/12 \times 1/12} = 0.4$$

c) denote these haplotypes as:  $\{a, b, c, d, e, f, g, h, i\}$

$$a = z_{21} / (3 \times 2) = 0.125 \quad b = z_{22} / (3 \times 2) = 0.042$$

$$c = z_{11} / (3 \times 2) = 0.067 \quad d = z_{12} / (3 \times 2) = 0.042$$

$$e = (z_{12} + z_{21} + z_{31}) / (3 \times 2) = 0.325 \quad f = z_{12} / (3 \times 2) = 0.1$$

$$g = (z_{11} + z_{12}) / (3 \times 2) = 0.133 \quad h = z_{32} / (3 \times 2) = 0.067$$

$$i = z_{31} / (3 \times 2) = 0.1$$

d)

Please see code for implementation.

```
>>>Frequency:
[ 0.167  0.    0.    0.    0.5   0.167  0.    0.    0.167]
Probability:
0.0,1.0
1.0,0.0
1.0,0.0
```

Haplotypes and corresponding frequencies are :

00010: 0.166

10011: 0.5

10101: 0.166

11111: 0.166