

# 1 Results of test cases

## 1.1 Test case\_1:

- Take a point estimate of theta based on the training data.
- For each amino acid, find the class that maximizes the probability that the resulting peptide will be a hit, and take an amino acid from this class.
- This constructs a peptide that approximately maximizes the probability of a hit(call it **app\_best**).
- Calculate the probability that it is hit (using simulation), and graph this probability vs. the prior (on a log -scale from  $10^{-4}$  to  $1/2$ ).

Figure 1 is the result when we set the length of app\_best to be 12(with 3 at the left of serene and 9 right):

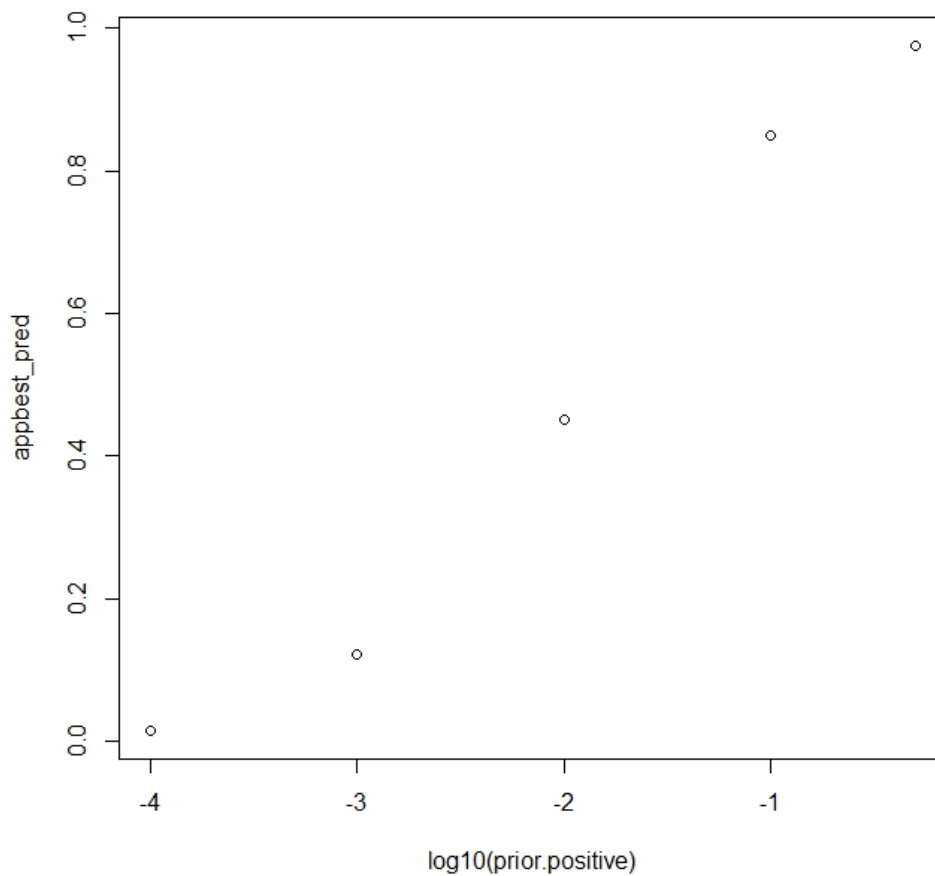


Figure 1: Result of Test Case\_1(Length = 12).

The length of `app_best` has a quite big influence on its probability being a hit, if we change the length to be 38(19 at each side of the serene), the result is shown in Figure 2.

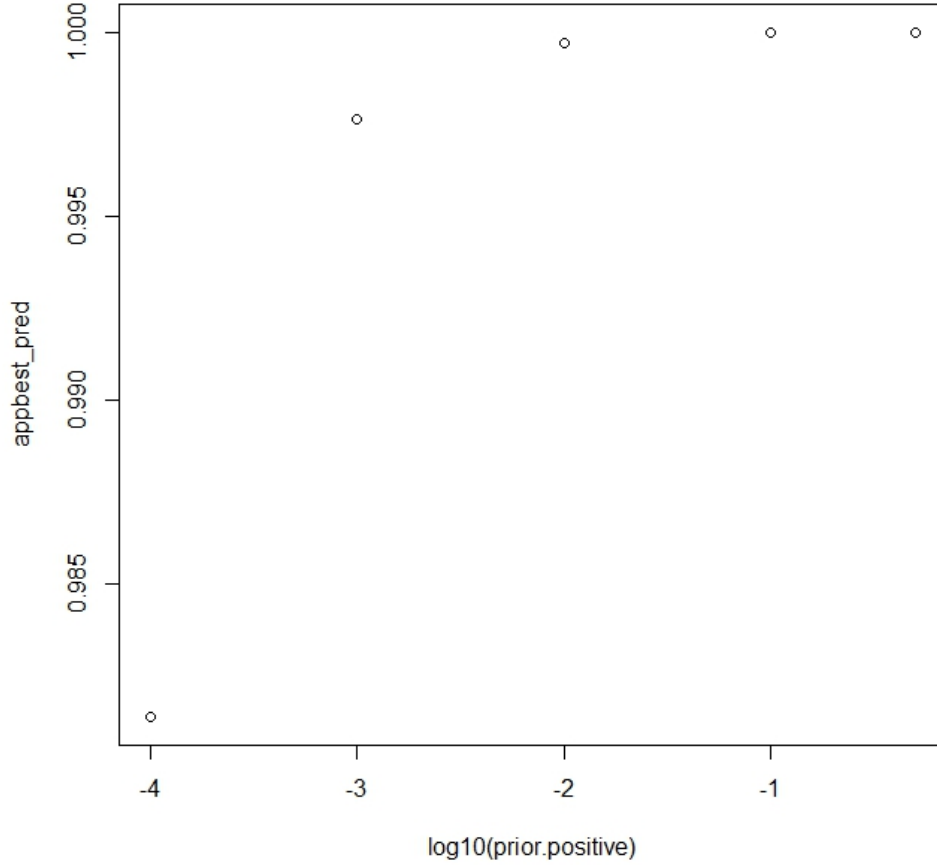


Figure 2: Result of Test Case\_1(Length = 38).

However, in either case, the probability is quite low when prior is  $10^{-4}$ .

## 1.2 Test Case\_2:

- Pick a pair of  $\theta$  (doesn't really matter how). Call these the true  $\theta$ .
- Choose a prior, try values in the range from  $10^{-4}$  to  $1/2$ .
- Generate training data ( $10^4$  peptides worth) where hit vs. no-hit is generated based on the true  $\theta$ .
- Generate a few other peptides ( $10^2$ ) & for each of these peptides, calculate:
  - Probability it is a hit based on the true  $\theta$ .

- Probability it is a hit based on the estimation method using simulation.
- For each peptide, plot a point which has these two probabilities on each axis.
- They should lie along a line.

Figure 3 is the result:

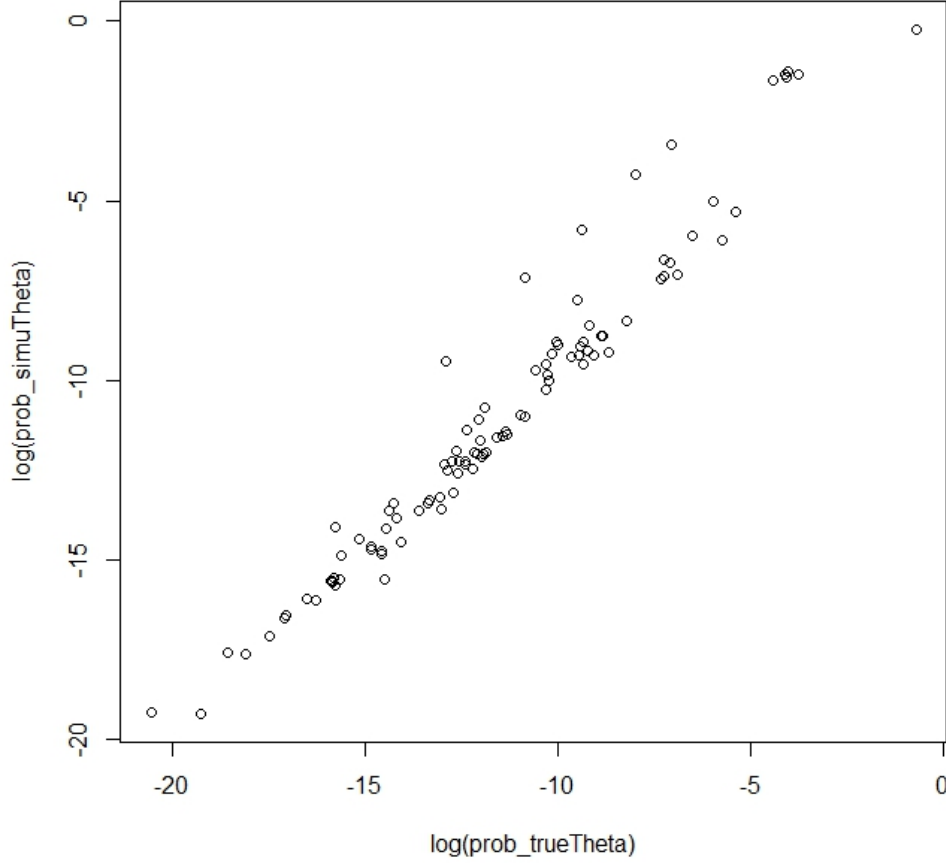


Figure 3: Result of Test Case\_2(Length = 12).

They do lie along a line.

## 2 Quality of recommending list

Figure 4 is the histogram of probability being hits of each peptide(total number is 300). The probability is computed using simulation(prior =  $10^{-4}$ ).

The average probability of the 300 peptides is: 0.013.

The average probability of the 300 peptides vs. prior(from  $10^{-4}$  to  $1/2$ ) is shown in Figure 5.

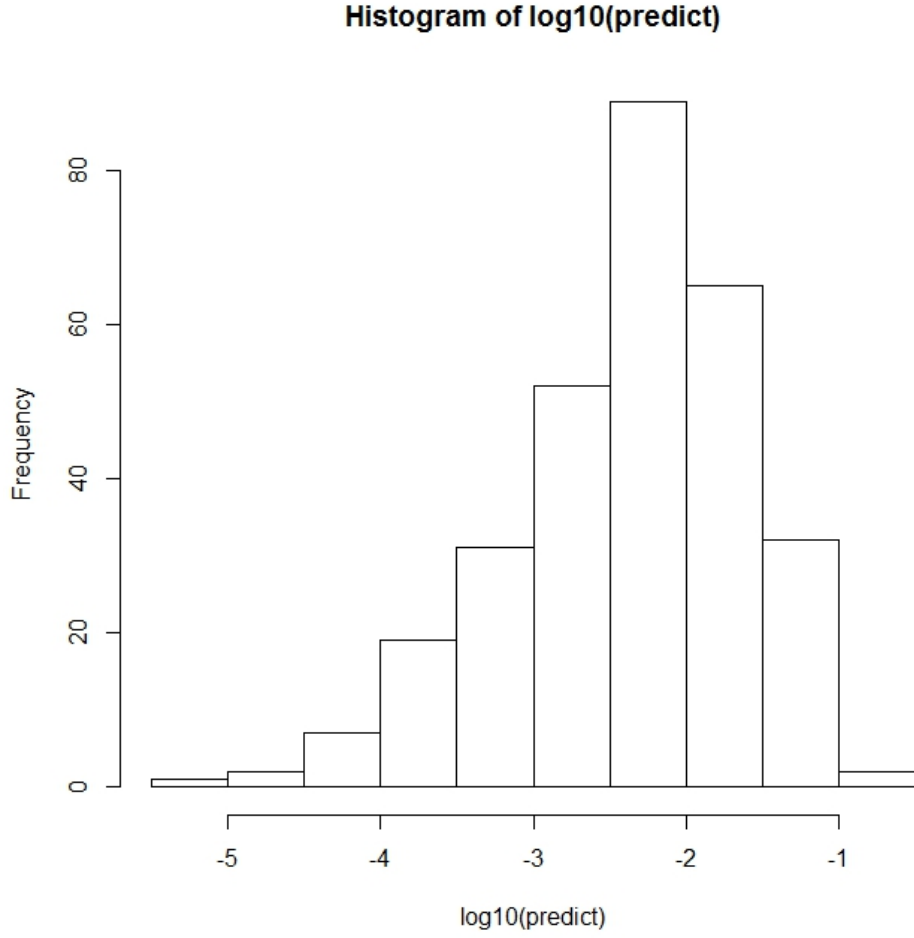
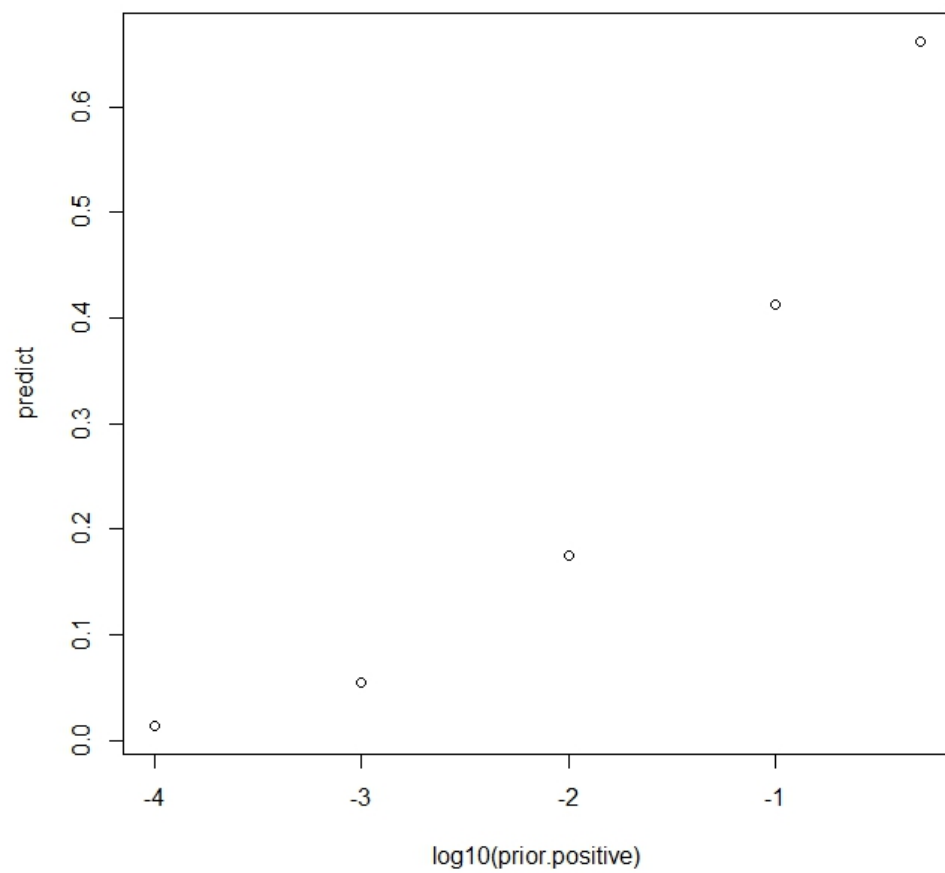


Figure 4: Histogram of probability of being hits

### 3 ProbImprovement & ExpImprovement

The probability that at least one peptide in our recommendation list is a hit is (computed using simulation, 100 replications): 0.82, with a 95% confidence interval [0.744, 0.895].

The expected reduction of length is 3.43, with a 95% confidence interval [2.993, 3.867].



prob.jpeg

Figure 5: prior vs. probability of being hits