Xiheng He

Lisanne Friedrich

Exercises for Algorithmic Bioinformatics II

Assignment 10

Xiheng He

Januar 2022

(Bayesian Modeling and p-values, 10P):

Modern Big-data based systems make all possible statements about behavior, persons and pre- ferences (voting behavior, buying or movie preferences etc.). Let's assume Facebook (or Google ("Google knows you"), Amazon, Apple, IBM, or Cambridge Analytics, the NSA, the FBI, the Verfassungsschutz, the BND, ...) develops a system to predict whether a person has a dangerous infectious disease X. Assume further that data on all 80 million inhabitants in Germany are available, the algorithm is very accurate and the program predicts that 100K people have a ten times higher probability of having X than the average person. From the press and offcials you know that currently there are about 10,000 infected people in Germany (i.e. relatively few, 1:8,000 or 0.0125%).

(a) If your neighbor is on the list, what is the probability that he/she is infected by X? Would you report him? Watch him skeptically or avoid him? Ignore the information? What if a family member is on the list?

P(infected by x | on the list) = 0.0125% × 10 = 0.125%

The probability of the neigbour being infected X is 0.125%

I would report him because X is a dangerousinfectious disease. Avoid him. Don't ingnore the information. Isolate the infected family number.

(b) Assess the situation in detail. Draw up a four-field table (Infected by X × Facebook list L) and describe the data situation.

data situation	Infected by X	Not infected by X
Facebook list L	0.125%	99.875%
Not on Facebook list L	0.0125%	99.9875%

(c) Define and compute the marginal/joint and conditional probabilities ($P(X), P(L), P(X,L), ..., P(-X|L), ...$).

$$P(X) = P(X|L) + P(X|\overline{L}) = 0.125\% + 0.0125\% = 0.1375\%$$

$$P(L) = 100000/800000000 = 0.125\%$$

$$P(X,L) = P(X|L)P(L) = 0.1375\% \times 0.125\% = 0.0171875\%$$

$$P(-X|L) = 1 - P(X|L) = 1 - 0.125\% = 99.875\%$$

(d) Discuss all probabilities, e.g., "What is the chance of ending up on the list?", "What is the probability of being on the list but not having X?", What is the probability (or degree of belief) that your family member is infected by X?

P(ending up on the list) $= \dfrac{100000}{800000000} = 0.125\%$

P(being on the list $\wedge$ not having x) = P(being on the list)P(not having x|being on the list) =

$0.125\% \cdot 99.875\% = 12.484375\%$

P(family member being infected by x) = P(family member being infected by x | not on the list) + P(family member being infected by x | on the list) = $0.125\% + 0.0125\% = 0.1375\%$

(e) Consider different relevant null hypotheses (e.g. "My friend is NOT infected by X!") and calculate p-values!

$P(X) = 0.1375\% \Longrightarrow K = 800, \lambda = 1$

"My friend is NOT infected by X is not a resonable null hypothesis" because we can't reject due to its unsignificant probability.

H_0: my friend is infected by X, H_1: my friend is not infected by X

The probability of being infected by X in period t for possion random variable:

$$P(X = k) = e^{-\lambda}\frac{(\lambda)^k}{k!}$$
$$p = Pr(X > 1) = \sum_{k=1}^{800} P(X = k) = \sum_{k=1}^{800} e^{-\lambda}\frac{(\lambda)^k}{k!} \approx 0 < 0.05$$

So we can reject the null hypothesis due to $p < 0.05$.