STAT 391
Homework 2
Out April 12, 2022
Due April 19, 2022

©Marina Meilă
mmp@cs.washington.edu

[**Problem 1 – Repeated sampling – NOT Grade**]

A man claims to have extrasensory perception. As a test, a fair coin is flipped 10 times, and the man is asked to predict the outcome in advance. He gets 7 out of 10 correct.

a. Let y refer to the number of correct tests, and denote the outcomes of the 10 individual tests with the random variables $X^1, X^2, ..., X^{10}$. What are the distributions of each X^i? What is the relationship between $X^{1:10}$ and Y? (No proof required).

b. What is the probability that he would have done at least this well if he had no ESP, i.e. if his guesses were essentially random?

c. Suppose the test is changed – now, the coin is flipped until the man makes an incorrect guess. He guesses the first two correctly, but guesses the third wrong. What is the probability of this experimental outcome (again, assuming no ESP)?

d. Assume both tests were planned beforehand. What is the probability that both of these tests turned out the way they did? In other words, the plan was to first flip the coin 10 times and count how many times the man is correct (Y) then to continue flipping until the man makes the next mistake, at flip $10 + Z$. You are asked the probability that $Y = 7$ and $Z = 3$.

Problem 2 – ML estimation with two models

Show your work. However, everything in the book or in class can be used without proof. E.g. you can use any ML estimation formulas from the lecture or book without proving them.

a. The Nisqually.com company sells books A,B,C on line. Each customer buys 0 or 1 copy of each title. Last week the company had 10 customers visit their online store. This is what the customers ordered:

A	B	C
1	0	0
1	0	0
1	0	0
0	1	0
0	1	0
1	0	0
0	0	1
0	0	1
0	0	0
1	0	0

Estimate $\theta_A = P_A(1), \theta_B = P_B(1), \theta_C = P_C(1)$ the probabilities that a customer orders books A, B, C respectively by the Maximum Likelihood method. We assume that customers' decision to buy each book is independent of the decision to buy other books, i.e.

$$P_{ABC}(x_A, x_B, x_C) = P_A(x_A)P_B(x_B)P_C(x_C) \quad (1)$$

b. What is the log-likelihood $l(\theta_A^{ML}, \theta_B^{ML}, \theta_C^{ML})$ of the data set above under the ML parameters estimated in **a**? Numeric answer only.

c. Now we assume another (equally simplistic) customer model. Namely, that each customer buys only one book, either A,B, or C. Customers who buy nothing are not counted. This models is represented by the probability distribution $\tilde{P} = (\tilde{\theta}_A, \tilde{\theta}_B, \tilde{\theta}_C)$ over $\tilde{S} = \{A, B, C\}$ with $\tilde{\theta}_A + \tilde{\theta}_B + \tilde{\theta}_C = 1$, $\tilde{\theta}_{A,B,C} \geq 0$.

The data observed from $\tilde{n} = 9$ customers is $A, A, A, B, B, A, C, C, A$ (note that this is the same data as in **a**, excluding the customer who buys nothing.

Estimate the parameters $\tilde{\theta}_{A,B,C}$ by the ML method.

d. What is the log-likelihood $\tilde{l}(\tilde{\theta}_A^{ML}, \tilde{\theta}_B^{ML}, \tilde{\theta}_C^{ML})$ of the data set above under the ML parameters estimated in **c**? Numeric answer only.

This exercise shows that the same data can be represented by different probabilistic models with different outcome spaces. Think whether the likelihoods l and $\tilde{l}$ are comparable. The two models impose different restrictions on the data. Incidentally, this toy data set satisfies the more restrictive second model.

The first model, when it is applied to natural language, is called the Bag of Words model, while the second model is called the Multinomial model.

[**e. Not graded**] Can you tell if the toy Language Classification experiment uses the Bag of Words model, the Multinomial, or neither?

Problem 3 – ML estimation with tied parameters

Zhenman rolls a die n times, and observes a data set $\mathcal{D}$ with counts $n_1, \dots n_6$. She is told that the die is not a fair one: the odd faces have the same probabilitly of coming up, denoted by θ_o, the even faces also have the same probabilitly of coming up, denoted by θ_e, but $\theta_o \neq \theta_e$, i.e. the distribution P defined by the die is given by $\theta_1 = \theta_3 = \theta_5 = \theta_o$ and $\theta_2 = \theta_4 = \theta_6 = \theta_e$.

a. Write the expression of the probability $P(3, 2, 1, 1, 6)$. What are the counts $n_{1:6}$ for this sequence of die rolls?

b. Write the expression of $l(\theta_o, \theta_e)$ the log-likelihood of the data $(3, 2, 1, 1, 6)$ as a function of θ_o, θ_e and the counts $n_{1:6}$.

c. Now consider an arbitrary data set $\mathcal{D}$ of size n, with counts $n_{1:6}$. Write the expression of $l(\theta_o, \theta_e)$ for this general case as a function of the counts $n_{1:6}$.

d. Transform $l(\theta_o, \theta_e)$ into a function of one variable, $l(\theta_e)$.

e. Now find the ML estimate of θ_e by equating the derivative of $l(\theta_e)$ with 0.

[**f–Extra credit**] Explain why the result above is intuitive/not surprising/natural.

Problem 4 – ML estimation with censored data

You record n samples from a Poisson distribution with rate parameter λ. However, due to a really poor choice of variable (namely, boolean), all that ends up being recorded is whether each sample was zero or non-zero. *In statistics, this is called an example of censored data.* Using only this information, and your knowlege of maximum likelihood estimation, what is your estimate of the rate parameter λ?

Problem 5 – The ML estimate as a random variable

Submit the code used for this problem

Consider the coin toss experiment ($m = 2$) with $\theta_1 = 0.3141$. The coin is tossed $n = 100$ times, obtaining independent outcomes from which we estimate the parameters $\theta_1^{ML}, \theta_0^{ML} = 1 - \theta_1^{ML}$ by the max likelihood method.

1. What is the set of possible values S_{θ_1} for θ_1^{ML} ? Does the true θ_1 belong to S_{θ_1}?

2. Write the expression of the probability of each outcome in S_{θ_1}, i.e the probability that $\theta_1^{ML} = j/n$ for $j = 0, 1, \dots n$.

3. Make a plot of the probability distribution of θ_1^{ML}. Preferably, this should be a "stem and flower" plot (the `stem` function in Matlab or python) like in figure 4.2 in the book. To avoid numerical overflow/underflow in the computation of the probabilities, consider using logarithms for the intermediate computations. The final results should not be in logarithm form, however. Take figure 4.2 as an example of how your plot should look like.

4. Let $\epsilon = 0.02$. Answer using the probability distribution computed previously (numerical answer only is OK):

$$\begin{aligned} \delta_{abs} &= P[|\theta_1^{ML} - \theta_1| > \epsilon] \ = ? \\ \delta_{rel} &= P[\frac{|\theta_1^{ML} - \theta_1|}{\theta_1} > \epsilon] \ = ? \end{aligned}$$