Amirkabir University of Technology
(Tehran Polytechnic)

Assignment 3
Estimation: From Parameters to Densities

Homeworks Guidelines and Policies

- ***What you must hand in.*** It is expected that the students submit an assignment report (.pdf) as well as – if necessary – required source codes (.m or .py) into an archive file named according to the following template: HW3_XXXXXXXX.zip
  where XXXXXXXX must be replaced by their student ID.
- ***Pay attention to problem types.*** Some problems are required to be solved *by hand* (shown by the icon), and some need to be implemented (shown by the icon).
  Please don't use implementation tools when it is asked to solve the problem by hand, otherwise you'll be penalized and lose some points.
- ***Don't bother typing!*** You are free to solve by-hand problems on a paper and include picture of them in your report. Here, cleanness and readability are of high importance. Images should also have appropriate quality.
- ***Reports are critical.*** Your work will be evaluated mostly by the quality of your report. Don't forget to explain what you have done, and provide enough discussions when it's needed.
- ***Appearance matters!*** In each homework, 5 points (out of a possible 100) belongs to compactness, expressiveness and neatness of your report and codes.
- ***Python is also allowable.*** By default, we assume you implement your codes in MATLAB. If you're using Python, you have to use equivalent functions when it is asked to use specific MATLAB functions.
- ***Be neat and tidy!*** Your codes must be separated for each question, and for each part. For example, you have to create a separate .m file for part b. of question 3. Please name it like p3b.m.
- ***Use bonus points to improve your score.*** Problems with bonus points are marked by the icon. These problems usually include uncovered related topics or those that are only mentioned briefly in the class.
- ***Moodle access is essential.*** Make sure you have access to Moodle because that's where all assignments as well as course announcements are posted on. Homework submissions are also done through Moodle.

- ***Assignment Deadline.*** Please submit your work **before the end of December 28th**.
- ***Delay policy.*** During the semester, students are given 7 free late days which they can use them in their own ways. Afterwards there will be a 25% penalty for every late day, and no more than three late days will be accepted.
- ***Collaboration policy.*** We encourage students to work together, share their findings and utilize all the resources available. However you are not allowed to share codes/answers or use works from the past semesters. Violators will receive a zero for that particular problem.
- ***Any questions?*** If there is any question, please don't hesitate to contact us through the following email addresses: **ali.the.special@gmail.com** and **dalirani@aut.ac.ir**. You may also find us in pattern recognition and image processing lab, 3rd floor, CEIT building.

1. Parameter Estimation: Let's Warm-up! (12 Pts.)

Keywords: *Parameter Estimation, Estimator, Likelihood Function, Maximum Likelihood Estimation, Maximum A Posteriori (MAP) Estimation, Bayesian Estimation, Posterior Distribution, Prior Distribution*

Parameter Estimation is a branch of methods that involve using samples data to estimate parameters of a distribution. The techniques used for parameter estimation are called **Estimators**. Amongst different estimators, **Maximum Likelihood** and **Maximum A Posteriori** methods are more common. While the first focuses on maximising the **Likelihood Function** given a set of **Observations**, the second attempts to minimise the posterior expectation of what is called the **Utility Function**.

Let's practice parameter estimation with some routine problems. First, suppose n samples $X_1, X_2, \ldots, X_n$ are drawn independently from a distribution having the PDF

$$f(x) = \frac{x^3 e^{-x/\theta}}{6\theta^4}, \quad \text{for } x \geq 0$$

a. Calculate the MLE estimator for θ. Make sure to justify all steps.
b. Find the MLE estimate for θ obtained from the following observations: 6.8, 7.2, 4.7, 7.9, 9.5, 6.1.

Next, consider a bag with some number θ of blue balls and a single red one. The parameter θ is unknown, yet we are informed that it would be one of the values of $\Theta = \{0,1,2,3,4\}$. Every time we draw a ball at random from the bag, noting its color, then return it to the bag. We repeat this experiment over and over again. Let R be the number of red balls observed in N independent repetitions of the experiment. Assume in five repetitions, we find a red ball four times.

c. Calculate the likelihood function $f_5(R = 4 \mid \theta)$ for this data.
d. Find the MLE estimator $\hat{\theta} \in \Theta$ for $R = 4$.
e. Considering a uniform prior distribution that assumes probability $\xi(\theta) = 0.2$ on each of the five possible values of $\theta \in \Theta$, find the posterior distribution.

Finally, consider a Printed Circuit Board (PCB) called *Quixant* has times-to-failure T_i which are independent random variables described with the following PDF:

$$f(x \mid \lambda) = \lambda^2 x e^{-\lambda x}, \quad x > 0$$

where $\lambda > 0$ are unknown and $1 \leq i \leq n$. We have conducted $n = 5$ experiments, and failure-times $T_1 = 3$, $T_2 = 5$, $T_3 = 1$, $T_4 = 4$ and $T_5 = 7$ are recorded.

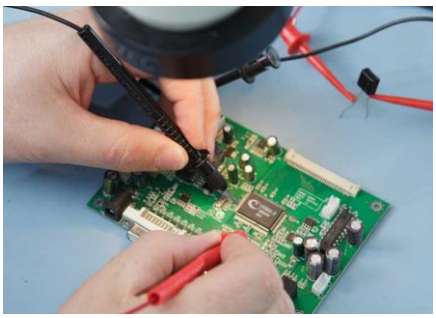


Figure 1 A prediction of reliability is an important element in the process of designing printed circuit boards (PCBs)

f. Obtain the Likelihood function $f_n(x \mid \lambda)$ for these observations.
g. Find the MLE estimate $\hat{\lambda}_n(x)$.
h. Assuming a uniform prior $\xi(\lambda) \equiv 1$, calculate the posterior density function $\xi(\lambda \mid x)$. The name of the distribution and the value(s) of the parameter(s) is also acceptable.
i. Write down the posterior mean $\mathrm{E}[\lambda \mid x]$ for this prior.

2. Don't Underestimate It: Parameter Estimation Can Save Lives! (15 Pts.)

Keywords: *Parameter Estimation, Maximum Likelihood Estimation, Bayes Estimation, Biased Estimator, Estimator Variance, Mean Squared Error (MSE)*

Silicon dioxide, aka *silica*, is an oxide of silicon which is used in various applications, including structural materials, microelectronics and pharmaceutical industries. However, inhaling crystalline silica dust may lead to serious diseases such as bronchitis and lung cancer. The European Environment Agency (EEA) conducts occasional reviews on its standards for airborne silica. During a review, the EEA investigates data from several studies. Each study takes into account different groups of people, and different groups have different exposures to silica.

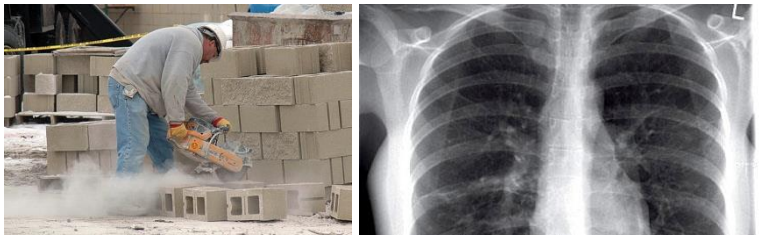


Figure 2 When people breathe silica dust, they inhale tiny little particles of the mineral silica, which can cause scar tissue in the lungs and even lead to lung cancer

Let s be the number studies, n_i be the number of people in the i'th study, x_i be the silica exposure for people in that study, and y_i be the number of people who developed lung cancer in that study. The EEA's model is $Y_i \sim \text{Poisson}(\lambda_i)$, where $\lambda_i = n_i x_i \lambda$ and λ is the typical rate at which silica causes cancer. The n_i's and x_i's are known constants, yet the Y_i's are random variables. Since different studies involve different groups of people in different places, they model the Y_i's from different studies as being independent, but not identically distributed since the λ_i's are different. The EEA goal is to estimate λ.

a. Write down the PDF of the joint distribution of $Y_1, \ldots, Y_s \mid \lambda$, which will also involve the constants $x_1, n_1, \ldots, x_s, n_s$.
b. Calculate the maximum likelihood estimator of λ.
c. Justify whether the maximum likelihood estimator is an unbiased estimator of λ.
d. Calculate the variance of the maximum likelihood estimator.
e. Calculate the mean squared error of the maximum likelihood estimator.
f. Assume the EEA attempts to estimate λ by using this model and combining data from 3 studies with data recorded in the table below. Write down an expression for the maximum likelihood estimate of λ.
   Note: Your answer should only involve numbers, not symbols. However, you don't need to simplify your expression.

Study Number (i)	Sample Size (n_i)	Exposure Level (x_i)	Cancer Case Count (y_i)
1	10	0.3	1
2	25	0.2	3
3	100	0.5	15

g. Suppose the EEA analysts decide to adopt a prior of $\Lambda \sim \text{Gamma}(\alpha, \beta)$, where α and β are known constant they choose to reflect their prior knowledge about λ. Find the posterior distribution for Λ. You should arrive at specific form for the posterior distribution, with parameters involving α, β, $x_1, \ldots, x_s$, $n_1, \ldots, n_s$, and $y_1, \ldots, y_s$.

3. A Bit History - Here's Why Hitler Hated MLE So Bad (12 Pts.)

Keywords: *Parameter Estimation, Maximum Likelihood Estimation, Biased Estimator, Estimator Variance*

German Tank Problem is one of the most well-known **MLE** problems. In World War II, the Allies attempted to determine the total number of German tanks, known as *Deutsche Panzer*, by looking at the serial numbers of those they had captured or destroyed. The German were labeling their tanks with serial number, sequentially from 1 to N, where N was the total number of tanks they had.

Figure 3 During the World War II, Allied intelligence successfully managed to estimate the number of German tanks using statistical methods

Assume you are a British engineer who works for the UK army, and you are trying to give an estimate of the total number of German tanks to your commanding officer. Suppose you are given a sample of $n \le N$ serial numbers $X_1,\ldots,X_n$, sampled randomly without replacement.

a. The first captured tank has a serial number of 50. What is your estimate of the total number of tanks? Justify your answer.
   a1. The next captured tank has a serial number of 30. Now what is your estimate?
   a2. The next serial number is 60. How does your estimation change now?
   a3. The serial number of the next tank is 70. What is your estimate now?
   a4. The next is 5. What is your new estimate?
   Explain the procedure or algorithm you used above.
b. Find the MLE for the parameter N.
c. Calculate the estimated bias and variance of MLE estimator.
d. It is possible to show

$$\mathrm{E}\left[\max(X_1,\ldots,X_n)\right]=\frac{n(N+1)}{n+1}$$

   Use this to obtain an unbiased estimator for the total number of tanks based on $\max(X_1,\ldots,X_n)$. Can you interpret this estimator?

4. Introduction to Non-Parametric Density Estimation (12 Pts.)

Keywords: *Density Estimation, Non-Parametric Methods, Parzen Windows, K-Nearest Neighbors, Kernel Density Estimation*

In some cases, a data sample may not resemble a known probability distribution or cannot be easily made to fit the distribution. In this this case, parametric density estimation is not suitable, and an alternative method must be used which do not consider a common distribution. These methods are referred to as **Non-Parametric** methods. **Parzen Windows Method** (or **Kernel Density Estimation**) and **K-Nearest Neighbor Estimation** are two popular methods of non-parametric techniques.

Here, we are going to deal with the problem of density estimation in non-parametric form. First, given the set of samples $X=\{-8,-7,-5,-2,0,2,3,4,5,7\}$,

a. Find the Parzen window estimate of $P_i(x)$ using a rectangular window with $h_i=1/\sqrt{i}$. Sketch the results for $i=1,4,11$.

b. Suppose $h_i = h/\sqrt{i}$, where h is a given constant. Discuss the shape of $P_i(x)$ for different choices of h.

Now, let's consider the following one-dimensional set of samples:

$Y = \{2, 3, 5, 6.5, 7.5, 8, 8.5, 9, 9.2, 9.5, 10, 10.8, 11.2, 11.3, 13, 15\}$.

c. Find the values of K-nearest neighbors estimate of $P_i(x)$ at $x=4$, $x=6$, $x=8$, $x=10$ and $x=12$. Set $k_i = \sqrt{i}$, where $i = 9, 16, 25$.

d. Sketch the estimated density.

Finally, assume $g(x)$ is an unknown density function and the following samples are drawn from it:

$X = \{2, 3, 5, 6, 6, 7, 8, 8, 8, 11, 12, 12, 14, 18, 20, 20\}$

e. Estimate the density at $y = 4, 9, 14$ using a Gaussian kernel $N(0,1)$.

5. Studying the Role of Smoothing Parameter in Kernel Density Estimation (12 Pts.)

Keywords: *Density Estimation, Non-Parametric Methods, Kernel Density Estimation, Smoothing Parameter (Bandwidth)*

In **Kernel Density Estimation**, the **Smoothing Parameter** or **Bandwidth** controls the number of samples or window of samples used to estimate the probability for a new point. While a large window may result in a coarse density with little details, a small window may also have too much detail and not be smooth or general enough to correctly cover new or unseen examples.

In this problem, we are to investigate the importance of this parameter in practice. Consider the following two probability density functions:

$$p_1(x) = \begin{cases} x & \text{if } 0 < x < 1 \\ 2 - x & \text{if } 1 < x < 2, \\ 0 & \text{otherwise} \end{cases} \qquad p_2(x) = xe^{-x^2/2}$$

a. Sketch the PDFs.

b. Generate N i.i.d samples from the given PDFs, assuming $N = \{10, 100, 1000\}$.

c. For a univariate Gaussian kernel, it is often recommended to select $h^* \approx 1.06\hat{\sigma}N^{-1/5}$, where h^* is the optimal choice of bandwidth, N is the number of samples and $\hat{\sigma}$ is the estimate of the standard deviation of the samples. Calculate the sample standard deviation, $\hat{\sigma}$. For each N, estimate the optimal value for bandwidth, $h^*(N)$.

d. Use kernel density estimation with a Gaussian kernel for each N to estimate the PDFs, considering three different bandwidth values $\{h^*(N)/3,\ h^*(N),\ 3*h^*(N)\}$.

e. Summarise your results by plotting the two PDFs estimates. For each of the given densities, you need to have 9 plots overall (18 in total). Overlay each plot with the groundtruth densities. Comment on the effects of h, N, and the kernel itself on the estimations you obtained.

Recommended MATLAB Functions: `rand()`, `subplot()`

6. Evaluating Non-Parametric Methods In Practice (12 Pts.)

Keywords: *Density Estimation, Non-Parametric Methods, Parzen Windows, K-Nearest Neighbors, Kernel Density Estimation, Kernel Function*

Amongst **Non-Parametric** methods for **Density Estimation**, some are more common. **Parzen Window** method, also known as **Parzen-Rosenblatt Window** method, utilises a **Kernel Function** and approximates the distribution of a given set of samples using a linear combinations of kernels centred on the observed points. **K-Nearest Neighbors** method, on the other hand, attempts to find the distance to the *k*-th nearest neighbors for a point *x* that we want to know the value of the distribution at. Obviously, these methods have their own pros and cons.

The aim of this problem is to compare these methods with respect to their performance on a certain dataset. Grab this one-dimensional dataset file "dense_data.dat" from the folder P6. Use the following non-parametric methods and the provided parameters to estimate the density associated with the given data. In each part, plot the resultant densities. Also, comment on the methods performance.

a. Parzen window with $h_1 = 0.0375$ and $h_2 = 0.075$.
b. K-nearest neighbors with $K_1 = 15$ and $K_2 = 30$.
c. Gaussian kernel density estimation with $\sigma_1 = 0.03$ and $\sigma_2 = 0.06$.

Note 1: Plot the data in the range of 0 to 1 at 0.001 intervals.

Note 2: Using built-in functions for density estimation are not allowed.

7. K-NN In Action (I): Regression Analysis (12 Pts.)

Keywords: *Regression Problem, K-Nearest Neighbors*

K-Nearest Neighbors algorithm is probably the simplest widely used model in machine learning. In spite of its simplicity, **K-NN** has proven to be incredibly powerful in various machine learning applications. Although it's far more popular in classification problems, K-NN can also be used in any regression task. The aim of this problem is to investigate how K-NN can be equally effective when the target variable is continuous in nature.

First, consider a simple regression problem with one dependent variable and one independent variable. Here, the goal is to predict an animal's body weight given its brain weight. You will use "animals_weight.txt" dataset which contains a list of brain weight and body weight measurements from a bunch of animals.

a. Given the following brain weights of some unknown animals, predict their body weight using 1-NN, 3-NN and 5-NN models.

#Test Sample	Brain Weight
1	53.298
2	1247.122
3	0.583
4	4.859
5	0.041

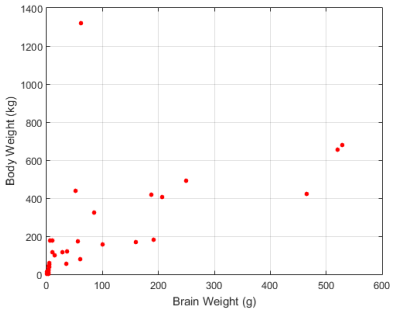


Figure 4 Distribution of brain and body weights of 64 species of animals, with two of the outlier samples removed for better visual purpose

b. Plot the regression line considering 1-NN, 3-NN and 5-NN models.

Now let's consider a more complicated problem, i.e. multivariate regression. You are given a dataset, "wine_quality.csv", containing 1600 red Vinho Verde wine samples. The goal is to model wine quality based on 11 physicochemical measurements.

Figure 5 The quality of red wine can somehow be evaluated by its physicochemical metrics such as the amount of citric acid or pH

c. Given the following table containing five set of physicochemical measurements, predict the corresponding wine qualities using 1-NN, 3-NN and 5-NN models.

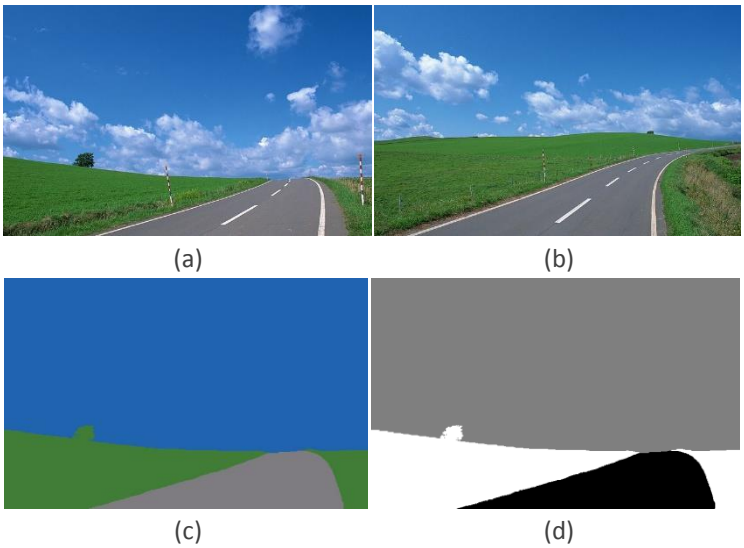
#	Fixed acidity	Volatile acidity	Citric acid	Residual sugar	chlorides	Free sulfur dioxide	Total sulfur dioxide	density	pH	suphates	alcohol	quality
1	7.5	0.9	0.26	2.3	0.054	19	13	0.99708	3.78	0.55	9.7	?
2	5.4	0.78	0.17	3.2	0.084	11	58	0.9987	2.94	0.83	11.8	?
3	8.2	0.56	0.46	1.7	0.069	25	15	0.997	3.39	0.65	12.5	?
4	6.0	0.7	0.01	4.6	0.093	6	104	0.99746	3.12	0.52	10.5	?
5	10.8	0.43	0.31	2.5	0.105	35	31	1.0001	3.22	0.48	11.1	?

8. K-NN In Action (II): Image Segmentation (12 Pts.)

Keywords: *Classification Problem, Image Segmentation, K-Nearest Neighbors*

Following previous problem, here we are going to evaluate **K-NN** classification capabilities in a different area. Imagine an **Image Segmentation** task where the goal is to detect and separate road regions of an input image. Such a system is necessary in many traffic-related applications, including lane departure warning system.

Here, two frames of a rural highway are available, Figure 6 part a and b. You are given a ground truth segmented version of frame a, which can somehow be used in the training phase of a K-NN model.



(a) (b)

(c) (d)

Figure 6 Two frames of a rural highway alongside the segmented version and segmentation mask of the first (a) frame a, which is used in training phase (b) frame b, which is used in test phase (c) segmented version of frame a (d) segmentation mask corresponding to frame a

a. Use 1-NN, 5-NN and 9-NN models to apply image segmentation on the image of the frame b. Display and compare the obtained results.
b. Do you think K-NN can be used in a lane detection system? Explain based on the observations you made in this problem.

Amirkabir University of Technology
(Tehran Polytechnic)

9. Some Explanatory Questions (8 Pts.)

Please answer the following questions as clear as possible:

a. Considering the fact that discrete PMF $f(x)$ is not differentiable with respect to x, is it possible to take a maximum likelihood estimate of a parameter for a discrete random variable?
b. When does MLE estimation lead to a better result than MAP estimation?
c. Why do we typically take a logarithm of the likelihood function?
d. Does K-NN algorithm have a training phase? Explain.
e. Suggest a method to determine a proper value for smoothing parameter (bandwidth) in Parzen window.
f. How would you interpret K-NN algorithm in terms of bias/variance?
g. Define an arbitrary two-category classification problem in 2D feature space where selecting different distance metrics affect K-NN results. More specifically, you must determine number of samples as well as their feature values.

Good Luck!
Ali Abbasi, Farhad Dalirani