STAT 391
Final Exam
10:30am – 12:20pm on June 8, 2023

©Marina Meilă
mmp@stat.washington.edu

Student Name:____

6 pages of notes are allowed + any solutions to homeworks

OK to write on backs of pages

No electronic devices of any kind are allowed during the exam.

Any fact that was proved in the lectures or in the notes can be used without proof.

Do Well!

Prob.1 ML with 2 models	of 5	______
Prob.2 Bias-Variance tradeoff for KDE	of 6	______
Prob.3 Likelihood ratio and Bayes classification	of 7	______
Prob.4 ML with censored data	of 6	______
Prob.5 Linear regression by ML	of 8	______
Bonus	1	
TOTAL:	of 33	______

(5 points) **Problem 1 – ML estimation with two models**

No need to show your work for this problem.

(1 point) *For the applicable problems you can provide either a numeric or symbolic answer.*

1.1 The Nisqually.com company sells books A,B,C on line. Each customer buys 0 or 1 copy of each title. We assume that customers' decision to buy each book is independent of the decision to buy other books, i.e.

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

What is the sample space S of the outcomes for one customer?

(1.5 points) **1.2** Last week the company had $n = 9$ customers visit their online store. This is what the customers ordered:

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

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 (ML) method.

(1.5 points) **1.3** Now we assume another (equally simplistic) customer model. Namely, that each customer buys only one book, either A,B, or C. 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 $\tilde{\theta}_A + \tilde{\theta}_B + \tilde{\theta}_C = 1$, $\tilde{\theta}_{A,B,C} \geq 0$.

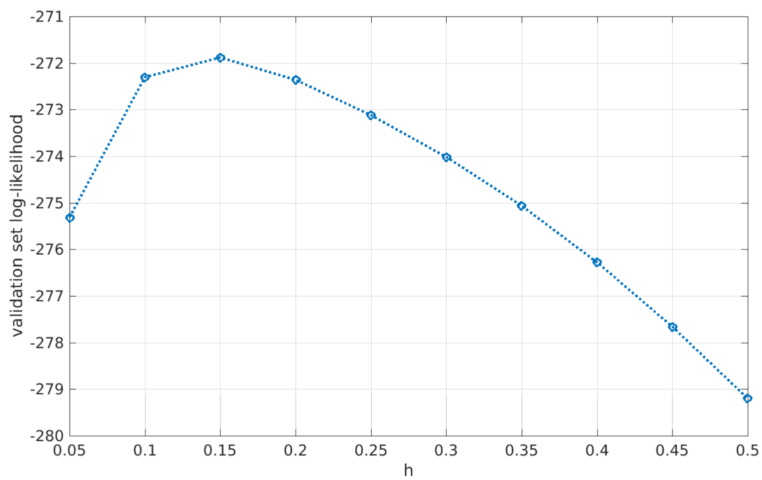
The data observed from $n = 9$ customers is $C, A, C, A, C, B, A, B, C$ (note that this is the same data as in **1.2**). Estimate the parameters $\tilde{\theta}_{A,B,C}$ by the ML method.

(1 point) **1.4** Give an example of an outcome that is in S but not in $\tilde{S}$.

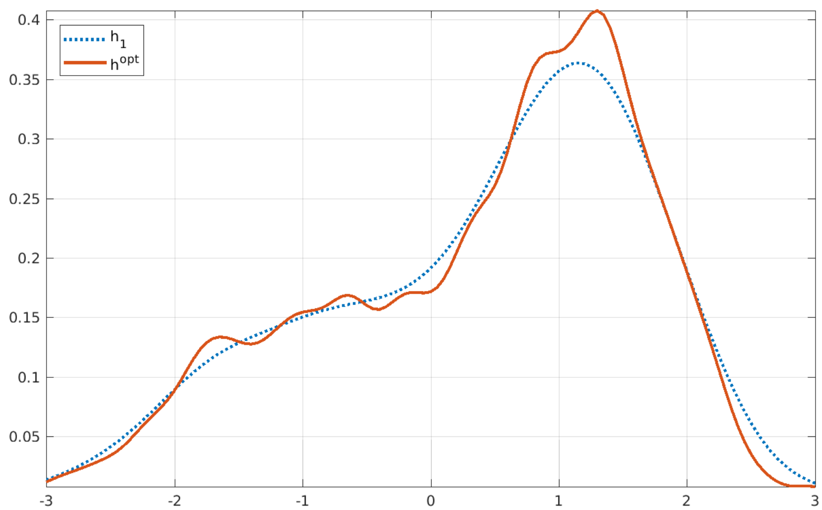
(6 points) **Problem 2 – Bias and Variance in Kernel Density estimation**

No proofs required for this problem

(1 point) **2.1** Assume that we have two data sets $\mathcal{D}, \mathcal{D}'$; we use the first for estimating a kernel density estimator $f_{h,K,\mathcal{D}}$, and the second a validation set. The graph below shows $l^{CV}(h; \mathcal{D}')$ the log-likelihood of $\mathcal{D}'$ under $f_{h,K,\mathcal{D}}$ (i.e. the CV log-likelihood), for different values of h. Based on this graph, what is h^{opt} the optimal kernel width value?



(5 points) **2.2** The graph below shows the kernel density estimators $f_{h^{\text{opt}},K,\mathcal{D}}$ and $f_{h_1,K,\mathcal{D}}$, with same data and kernel K for bandwidth values h^{opt} and $h_1 \neq h^{\text{opt}}$.

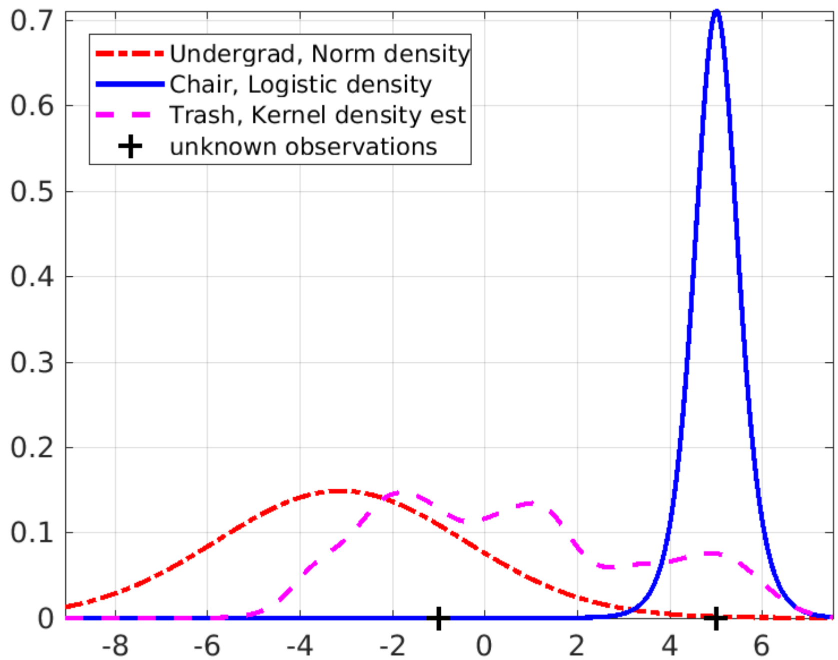


Answer based on the above graph.

$h_1 > h^{\text{opt}}$	TRUE	FALSE
$\text{Variance}(h_1) > \text{Variance}(h^{\text{opt}})$	TRUE	FALSE
$\text{Bias}(h_1) > \text{Bias}(h^{\text{opt}})$	TRUE	FALSE
$l^{CV}(h;\mathcal{D}') > l^{CV}(h^{\text{opt}};\mathcal{D}')$	TRUE	FALSE
$l(h;\mathcal{D}') > l(h^{\text{opt}};\mathcal{D}')$	TRUE	FALSE

(7 points) **Problem 3 – Likelihood Ratio and Bayesian classification**

The graph below shows three different densities on $(-\infty, \infty)$, the first "Undergrad"$= f_U$, is a normal density, the second, "Chair"$= f_C$ is a logistic density, the third "Trash"$= f_T$ is obtained by kernel density estimation. There are also 2 observations $x^1 = -1$, $x^2 = 5$. Each of x^1, x^2 was sampled from one of $f_{U,C,T}$ but not necessarily the same one.



(1 point) **3.1** We would like to know which of $f_{U,C,T}$ has the highest likelihood to have generated x^1, based on the graph. *Give a 1-line explanation of your answer.*

(1 point) **3.2** We would like to know which of $f_{U,C,T}$ has the highest likelihood to have generated x^2, based on the graph. *Give a 1-line explanation of your answer.*

(2.5 points) **3.3** Now we have some prior information. We know that the prior probabilities of f_U, f_C, f_T generating x^1 are respectively $P^0[f_U] = \pi_U = 1/2$, $P^0[f_C] = \pi_C = 1/3$, and $P^0[f_T] = \pi_T = 1/6$. Using this information, write the formulas for the posterior probabilities $P[f_U|x^1]$, $P[f_C|x^1]$, $P[f_T|x^1]$ of $f_{U,C,T}$ having generated x^1.

(1 point) **3.4** Do the above probabilities always sum to 1? TRUE FALSE

(1.5 points) **3.5** Using the graph and the information in **3.3**, which of $f_{U,C,T}$ is a-posteriori more probable to have generated x^1? *Give a 1-2 line explanation of your answer.*

(6 points) **Problem 4 – ML estimation with censored data**

Show your work

You are given samples $\{x^1, \dots x^n\}$ from a *geometric distribution* with unknown parameter γ:

$$P(x) = (1-\gamma)\gamma^x \quad \text{for } x \in \{0, 1, 2, \dots\}$$

But, by mistake, you store the data in the wrong format, which only preserves whether the data point was 0 or not.

$$y^i = \begin{cases} 0 & \text{if } x^i = 0 \\ 1 & \text{if } x^i \geq 1 \end{cases} \quad \text{for } i = 1:n. \tag{2}$$

We say that the y^i observations are *censored* observations of the data x^i. With only the censored data $\{y^1, \dots y^n\}$ you will estimate γ.

(1.5 points) **4.1** Write the probability that $y^i = 1$ as a function of γ.

(1.5 points) **4.2** Derive the expression of the log-likelihood $l(\gamma) = \ln P(y^{1:n}|\gamma)$ as a function of γ.

(1.5 points) **4.3** Maximize $l(\gamma)$ w.r.t. γ and obtain the expression for γ^{ML}.

(1.5 points) **4.4** Does this problem have sufficient statistics? How many and what are they?

(8 points) **Problem 5 – Linear regression by Maximum Likelihood**

Show your work

The data set $\mathcal{D} = \{(x^i, y^i),\ i = 1:n\}$ has $x^{1:n} \in [0,1]$ and y^i sampled as follows

$$y^i = \beta x^i + \varepsilon^i \quad \text{for } i = 1:n \tag{3}$$

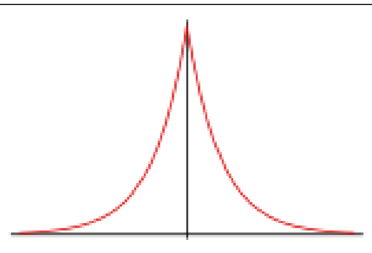
with

$$\varepsilon^i \sim \text{i.i.d., Laplace}(\gamma), \quad \text{for } i = 1:n \quad \text{with } f_{\text{Laplace}}(z) = \frac{\gamma}{2} e^{-\gamma|z|},\ z \in \mathbb{R}$$

We would like to estimate the parameters of the model in equation (3) by the Maximum Likelihood method.

(1 point) **5.1** What are the parameters of the model in equation (3)?

(1.5 points) **5.2** Write the expression of the likelihood $L(y^{1:n} \mid \beta, \gamma, x^{1:n})$; simplify it as much as possible.



Notice that the Laplace density is the exponential density, symmetrized around 0.

(1.5 points) **5.3** Write the expression of the log-likelihood $l(y^{1:n} \,|\, \beta, \gamma, x^{1:n})$; simplify it as much as possible.

[**5.4 Extra credit**] The ML method requires you to maximize over β, γ the expression of the log-likelihood. Show that $\beta^{ML} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{n} |y^i - \beta x^i|$. Is this a Least-Squares problem?

(2.5 points) **5.5** Assume that you have estimated β^{ML}. Denote $\hat{\varepsilon}^i = y^i - \beta^{ML} x^i$ (they are known, since β^{ML} is known). Derive the expression for γ^{ML} the ML estimator of γ; simplify it as much as possible.

(1.5 points) **5.6** Your data set has $n = 100$ samples ($\ln 10 = 2.30$). You have estimated β^{ML}, γ^{ML} and the log-likelihood $l(\mathcal{D}|\beta^{ML}, \gamma^{ML}) = -101.0$. You have also estimated another model from the same data, in which

$$y^i \ = \beta_0 + \beta_1 x^i + \tilde{\varepsilon}^i \quad \text{with } \tilde{\varepsilon}^i \sim \text{ i.i.d., Normal}(0, \sigma^2) \quad \text{for } i \ = \ 1 : n. \tag{4}$$

The log-likelihood of the model in equation (4), for the ML parameters, is $\tilde{l}(\mathcal{D}|\beta_0^{ML}, \beta_1^{ML}, (\sigma^2)^{ML}) = -100.1$. Select the best of these models, using AIC.

$$\boxed{\text{AIC(model)} \ = \ l(\mathcal{D}|\text{model}^{ML}) - \#\text{parameters(model)}}$$

[**5.7 Extra credit**] Show that $l(y^{1:n} \mid \beta^{ML}, \gamma^{ML}, x^{1:n})$ is independent of β^{ML}.

[extra space for anything]

Have a nice summer!