Midterm exam, 機器學習, Fall 2020. Open book but no calculators/cell phones allowed. Answers may include e^2, $\sqrt{2}$, etc. but simplify when possible.

Your Name: ____

Problem 1.
Suppose
1. we have some observed data X (a set of real numbers $\{x_1, ..., x_n\}$
2. We assume the data are random samples generated by a normal distribution of unknown mean μ.

Question 1**a**
What is the maximum likelihood estimator for the μ?

Question 1**b**
Given $\mu = 0$, what is the maximum likelihood estimator for σ^2?

Give the mathematical derivation for your answers:

Your Name: ____________________

Problem 2.
Let v be a random variable defined by these values and probabilities.

v	probability
2	0.4
3	0.3
4	0.2
5	0.1

Let $V(n) = v_1 + v_2 + \cdots + v_n$ be the sum of n independent samples of v.

Question 2
Derive the mean and standard deviation of $V(n)$.

Your Name: ______________________________

Problem 3.
The Poisson distribution has parameter $\lambda \geqq 0$, defining a probability distribution over the non-negative integers $(0, 1, ...)$ as follows:

$$\mathrm{Pois}(k; \lambda) \quad \stackrel{\text{def}}{=} \quad P[k] = \frac{\lambda^k}{k! \exp(\lambda)}, \quad k \in \mathbb{N}_0$$

This problem involves inference from data generated by one of two Poisson distributions: $\mathrm{Pois}(\lambda_1)$ or $\mathrm{Pois}(\lambda_2)$. The following experiment is done.
1. λ is set to $\{\lambda_1, \lambda_2\}$ with probability m_1 and $m_2 = 1 - m_1$.
2. A random sample y is drawn from $\mathrm{Pois}(\lambda)$

Question 3a
What is the posterior probability $\mathrm{P}[\lambda = \lambda_1 | y = k]$?

Question 3b:
What kind of prior is this? Is it conjugate? Why or why not?

Your Name: ____

Problem 4.
A standard poker deck has 52 cards. 13 each of: ♠ ♣ ♥ ♦ . The entropy of a single card drawn at random is ≈ 5.7 bits of information. You cannot see the card, but I can.

Question 4a
If I told you the card is black (i.e. '♠' or '♣'); how much entropy would remain? (give numerical answer and reason)

Question 4b
If I then told you the card was a spade '♠', how much entropy would remain then? (give numerical answer and reason)

Question 4c
Two cards are drawn from a fresh deck of cards. Let S_1, S_2 denote the first and second cards respectively. What is the mutual information $\mathrm{I}(S_1, S_2)$ (answer can include lg symbol).

Your Name: ______________________________

Problem 5. Consider a classification problem with two features $F1 \in \{0, 1, 2\}$, $F2 \in \{0, 1, 2, 3\}$, Assume we know the two classes occur with equal probability: $P[C = A] = P[C = B] = 0.5$ (so you do not need to estimate $P[C = A]$, just take it as given to be 0.5).

Training Data

F1	F2	Class
2	3	A
0	0	A
2	2	A
2	0	A
2	0	B
2	0	B
1	0	B
2	0	B

Test Data; P[C=A|F1,F2]:P[C=B|F1,F2] Using Prior:

F1	F2	MLE	Jeffreys	Laplace
2	3			
1	1			
0	3			
1	0			
0	0			
1	3			
0	2			
0	1			

The above table gives the P[F|C] probabilities for each feature and class.

Question 5

Compute the probability a Naïve Bayes classifier would assign to $P[C = A]$, using maximum estimation, Jeffrey's priors or Laplace priors respectively when estimating probabilities involving feature values. You may report the answer in terms of odds, so for example, if the $P[C = A] = \frac{1}{3}$, you can report that as 1:2 (hint: it is easier to work with odds).

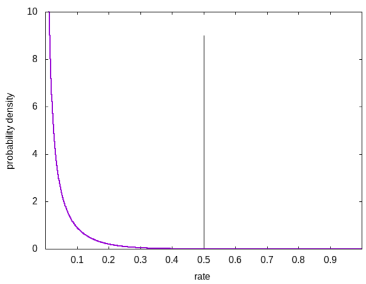
You may find the following worksheet helpful.

	Feature F1			Feature F2				Class
Value	**0**	**1**	**2**	**0**	**1**	**2**	**3**	
								A
counts								B
"counts"								A
Jeffreys								B
2×counts								A
Jeffreys								B
								A
Laplace								B

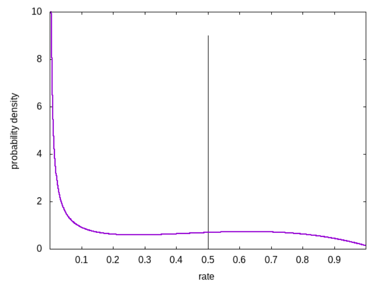
Your Name: ______________________________

Problem 6.

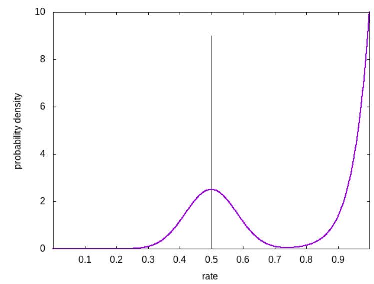
A



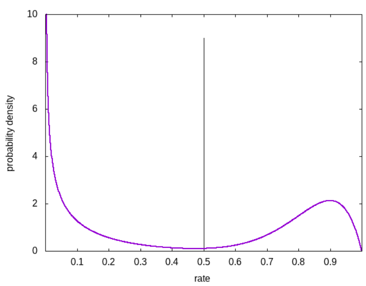
B



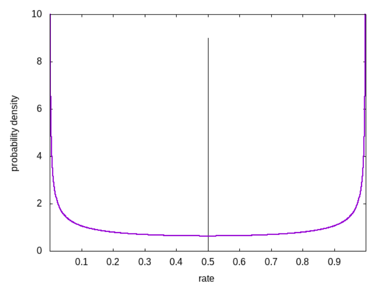
C



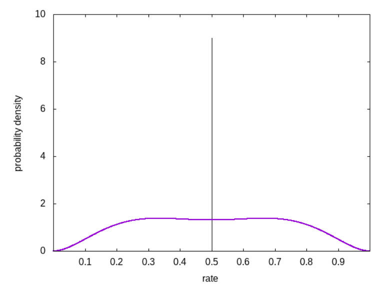
D



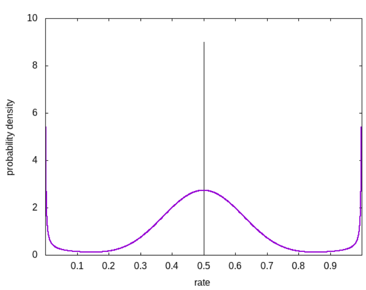
E



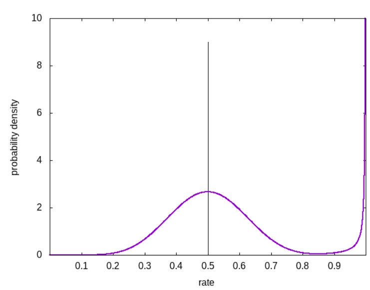
F



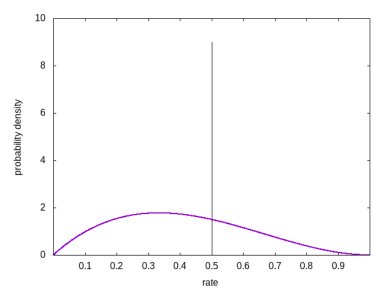
G



H



I



Question 6
What parameter values in the table on the previous path match which plot in the table above? Fill in the "Label" column and use the "Comment" column and/or space at bottom to explain your answers.

ID	Distribution	Label (A–I)	Comment
1	$0.8\,\mathrm{Beta}(9,9) + 0.2\,\mathrm{Beta}(9,0.1)$		
2	$\mathrm{Beta}(0.5,0.5)$		
3	$0.5\,\mathrm{Beta}(3,6) + 0.5\,\mathrm{Beta}(6,3)$		
4	$\mathrm{Beta}(2,3)$		
5	$0.5\,\mathrm{Beta}(20,20) + 0.5\,\mathrm{Beta}(20,1)$		
6	$0.3\,\mathrm{Beta}(3,2) + 0.7\,\mathrm{Beta}(0.2,1)$		
7	$\mathrm{Beta}(0.2,9)$		
8	$0.5\,\mathrm{Beta}(0.5,5) + 0.5\,\mathrm{Beta}(10,2)$		
9	$0.8\,\mathrm{Beta}(9,9) + 0.2\,\mathrm{Beta}(0.2,0.2)$		