Graded Assignment 4

Due date for this assignment: 2024-10-20, 23:59 IST.

You may submit any number of times before the due date. The final submission will be considered for grading.

For all questions involving the Bernoulli distribution, the parameter p is $P(x = 1)$.

1) Consider a dataset that has 10 zeros and 5 ones. What is the likelihood function if we assume a Bernoulli distribution with parameter p as the probabilistic model?

- ○ p^{15}
- ○ $(1-p)^{15}$
- ○ $p^{10} \cdot (1-p)^{5}$
- ● $p^{5} \cdot (1-p)^{10}$

2) In the previous question, what is the estimate of $\hat{p}_{ML}$? Enter your answer correct to two decimal places.

0.33

3) Consider a dataset that has a single feature (x). The first column in the table below represents the value of the feature, the second column represents the number of times it occurs in the dataset.

x	Frequency
-1	1
0	1
2	4
4	2
5	2

If we use a Gaussian distribution to model this data, find the maximum likelihood estimate of the mean.

- ○ 2
- ○ 0
- ● 2.5
- ○ The mean cannot be computed as the variance of the Gaussian is not explicitly specified.

4) Consider a beta prior for the parameter p of a Bernoulli distribution:

$p \sim \text{Beta}(3, 2)$

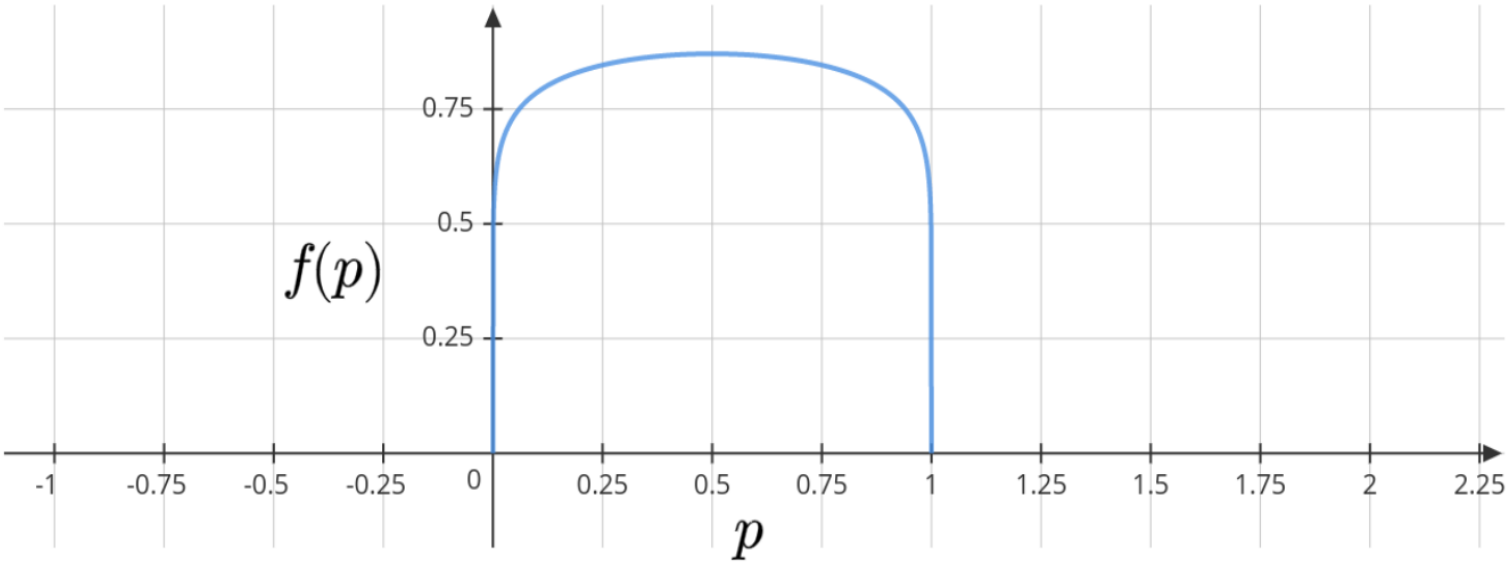
The dataset has 15 ones and 10 zeros. What is the posterior?

- ○ $\text{Beta}(3, 2)$
- ○ $\text{Beta}(13, 17)$
- ● $\text{Beta}(18, 12)$
- ○ $\text{Beta}(17, 11)$

5) In the previous question, we use the expected value of the posterior as a point-estimate for the parameter of the Bernoulli distribution. What is $\hat{p}$? Enter your answer correct to two decimal places.

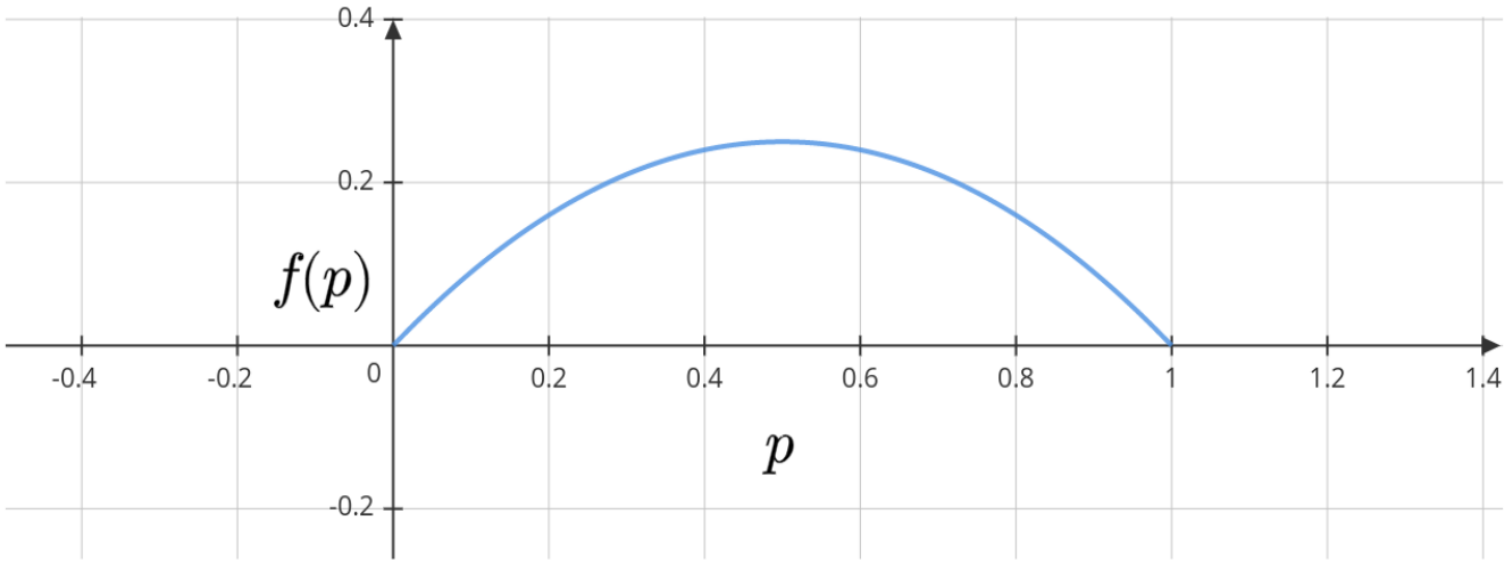
0·60

6) Consider the following prior distribution (Beta) of the parameter p of a Bernoulli distribution: **1 poi**



After observing 10 data-points, the following is the posterior distribution:

After observing 10 data-points, the following is the posterior distribution:



Ignore the values on the Y-axis and just focus on the shapes of the distributions. Which of the following could correspond to the observed data?

- ○ $\{1, 1, 1, 0, 1, 1, 0, 1, 1, 1\}$
- ○ $\{0, 1, 0, 0, 0, 1, 0, 0, 0, 0\}$
- ● $\{1, 1, 0, 1, 0, 0, 0, 1, 1, 0\}$

Common Data for questions (7) to (9)

We wish to fit a GMM with $K = 2$ for a dataset having 4 points. At the beginning of the t^{th} time step of the EM algorithm, we have $\theta^{(t)}$ as follows:

$\pi_1 = 0.3, \quad \pi_2 = 0.7$
$\mu_1 = 2, \quad \sigma_1^2 = 1$
$\mu_2 = 3, \quad \sigma_2^2 = 1$

The density of the points given a particular mixture is given to you for all four points. f is the density of a Gaussian.

x_i	$f(x_i \mid z_i = 1)$	$f(x_i \mid z_i = 2)$
1	0.242	0.054
2	0.399	0.242
3	0.242	0.399
4	0.054	0.242

Use three decimal places for all quantities throughout the questions.

7) What is the value of λ_k^i for $i = 1$ and $k = 2$ after the E-step? Enter your answer correct to three decimal places.

0.342

8) If we pause the algorithm at this stage (after the E-step) and use the λ_k^i values to do a hard-clustering, what would be the cluster assignment? We use the following rule to come up with cluster assignments:

$z_i = \underset{k}{\operatorname{argmax}}\, \lambda_k^i$

The answer is in the form of a vector: $\mathbf{z} = \begin{bmatrix} z_1 & z_2 & z_3 & z_4 \end{bmatrix}^T$.

- ○ $\begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}^T$
- ○ $\begin{bmatrix} 2 & 2 & 2 & 2 \end{bmatrix}^T$
- ○ $\begin{bmatrix} 1 & 1 & 2 & 2 \end{bmatrix}^T$
- ● $\begin{bmatrix} 1 & 2 & 2 & 2 \end{bmatrix}^T$

9) What is the value of μ_1 after the M-step? Enter your answer correct to three decimal places.

1·900

10) A GMM is fit for a dataset with 5 points. At some time-step in the EM algorithm, the following are the values of λ_k^i for all points in the dataset for the k^{th} mixture after the E-step:

$\lambda_k^1 = 0.3$
$\lambda_k^2 = 0.1$
$\lambda_k^3 = 0.4$
$\lambda_k^4 = 0.8$
$\lambda_k^5 = 0.2$

What is the estimate of π_k after the M-step? Enter your answer correct to two decimal places.

0·36

11) What is the value of the following expression after the E-step at time-step t in the EM algorithm? There are 100 data-points and 3 mixtures.

$$\sum_{i=1}^{100} \sum_{k=1}^{3} \lambda_k^i$$

- ○ 3
- ● 100
- ○ 103
- ○ 300
- ○ 1
- ○ The answer depends on the time-step t we are at