Graded

This document has 11 questions.

Note to Learners

Statement

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

Question-1

Statement

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?

Options

(a)

$$p^{15}$$

(b)

$$(1-p)^{15}$$

(c)

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

(d)

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

Answer

(d)

Solution

We shall use the i.i.d. assumption. If x_i is the random variable corresponding to the i^{th} data-point, we have:

$$P(x_i = 1) = p$$

x_i and x_j are independent for $i \neq j$ and they are identically distributed. The likelihood is therefore the product of 15 terms, five of which correspond to ones and the rest to zeros:

$$L(p; D) = p^5 \cdot (1-p)^{10}$$

Question-2

Statement

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

Answer

0.33

Range: $[0.32, 0.34]$

Solution

The estimate is the fraction of ones:

$$\hat{p} = \frac{5}{15} = \frac{1}{3} \approx 0.33$$

Question-3

Statement

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.

Options

(a)

2

(b)

0

(c)

2.5

(d)

The mean cannot be computed as the variance of the Gaussian is not explicitly specified.

Answer

(c)

Solution

$$\hat{\mu}_{_{ML}} = \frac{-1 + 0 + 4 \times 2 + 2 \times 4 + 2 \times 5}{10} = 2.5$$

Question-4

Statement

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?

Options

(a)

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

(b)

$\text{Beta}(30, 17)$

(c)

$\text{Beta}(18, 12)$

(d)

$\text{Beta}(17, 11)$

Answer

(c)

Solution

Since the beta distribution is a conjugate prior of the Bernoulli distribution, the posterior is also a beta distribution. Specifically, if the prior is $\text{B}(\alpha, \beta)$ and the dataset has n_1 ones and n_0 zeros, then the posterior:

$$\text{Beta}(\alpha + n_1, \beta + n_0)$$

In this problem, $\alpha = 3, \beta = 2, n_1 = 15, n_0 = 10$.

Question-5

Statement

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.

Answer

0.6

Range: $[0.59, 0.61]$

Solution

The expected value of a beta distribution with parameters α and β is:

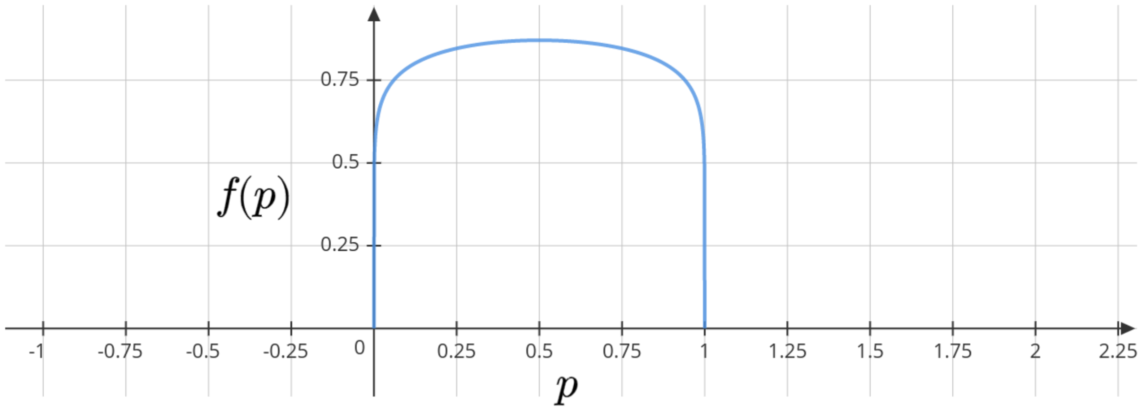
$$\frac{\alpha}{\alpha + \beta}$$

For the posterior, we have $\alpha = 18, \beta = 12$.

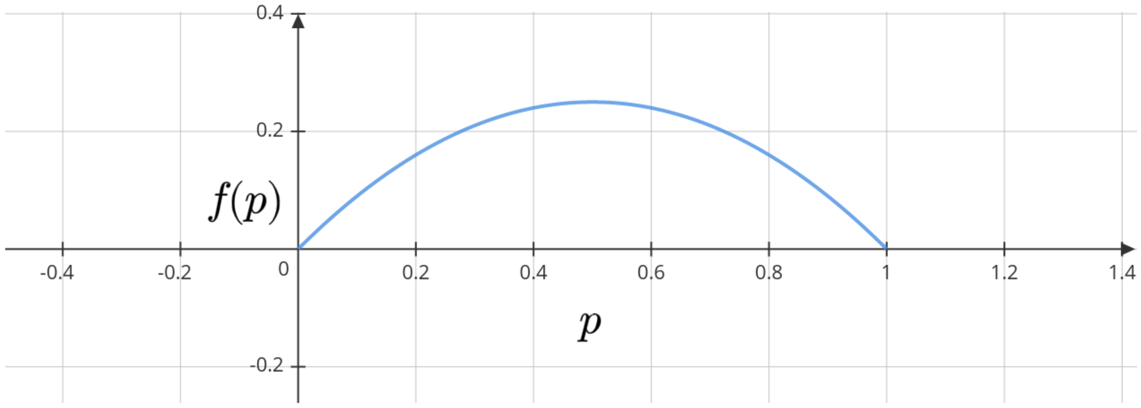
Question-6

Statement

Consider the following prior distribution (Beta) of the parameter p of a Bernoulli 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?

Options

(a)

$$\{1, \quad 1, \quad 1, \quad 0, \quad 1, \quad 1, \quad 0, \quad 1, \quad 1, \quad 1\}$$

(b)

$$\{0, \quad 1, \quad 0, \quad 0, \quad 0, \quad 1, \quad 0, \quad 0, \quad 0, \quad 0\}$$

(c)

$$\{1, \quad 1, \quad 0, \quad 1, \quad 0, \quad 0, \quad 0, \quad 1, \quad 1, \quad 0\}$$

Answer

(c)

Solution

The prior encodes the belief that coin is somewhat unbiased. The posterior seems to have made that belief stronger. So, the data should have been something that strengthens the belief in the prior, meaning, an equal number of ones and zeros.

Common Data for questions (7) to (9)

Statement

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.

Question-7

Statement

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.

Answer

0.342

Range: $[0.33, 35]$

Solution

From Bayes rule, we have:

$$\lambda_k^i = P(z_i = k \mid x_i) = \frac{P(z_i = k) \cdot f(x_i \mid z_i = k)}{f(x_i)}$$

So:

$$\lambda_2^1 = \frac{0.7 \times 0.054}{0.7 \times 0.054 + 0.3 \times 0.242} \approx 0.342$$

Question-8

Statement

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: $\mathbf{z} = \begin{bmatrix} z_1 & z_2 & z_3 & z_4 \end{bmatrix}^T$.

Options

(a)

$$\begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}^T$$

(b)

$$\begin{bmatrix} 2 & 2 & 2 & 2 \end{bmatrix}^T$$

(c)

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

(d)

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

Answer

(d)

Solution

We need to compute the table of λ_k^i values from which we can read of the cluster assignments.

x_i	λ_1^i	λ_2^i	z_i
1	0.658	0.342	1
2	0.414	0.586	2
3	0.206	0.794	2
4	0.087	0.912	2

Question-9

Statement

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

Answer

1.796

Range: $[1.7, 1.9]$

Solution

Now for μ_1:

$$\mu_1 = \frac{\sum\limits_{i=1}^{4} \lambda_1^i x_i}{\sum\limits_{i=1}^{4} \lambda_1^i}$$

Which is:

$$\mu_1 = \frac{0.658 \times 1 + 0.414 \times 2 + 0.206 \times 3 + 0.087 \times 4}{0.658 + 0.414 + 0.206 + 0.087} \approx 1.796$$

Question-10

Statement

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.

Answer

0.36

Range: $[0.35, 0.37]$

Solution

$$\pi_k = \frac{1}{5} \cdot \sum_{i=1}^{5} \lambda_k^i = \frac{0.3 + 0.1 + 0.4 + 0.8 + 0.2}{5} = \frac{1.8}{5} = 0.36$$

Question-11

Statement

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.

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

Options

(a)

3

(b)

100

(c)

103

(d)

300

(e)

1

(f)

The answer depends on the time-step t we are at

Answer

(b)

Solution

We know that the λ_k^i values should sum to 1 for each data-point. Since there are 100 data-points, the expression should sum to 100.