Week_4_mlt_questions

Jan 2024 Quiz 1

1. **Comprehension:**
Assume that you have a dataset of six points $\{x_1, x_2, x_3, x_4, x_5, x_6\}$, all of which are non-negative. You hypothesise that the data points are iid random variables with the following density:

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Based on the above data, answer the given subquestions.

i. What is the log-likelihood of this dataset under this distribution? ln represents the natural logarithm or $\log_e$.

(a) $\prod_{i=1}^{6} \lambda e^{-\lambda x_i}$

(b) $\sum_{i=1}^{6} \lambda e^{-\lambda x_i}$

(c) $\prod_{i=1}^{6} [\ln(\lambda) - \lambda x_i]$

(d) $\sum_{i=1}^{6} [\ln(\lambda) - \lambda x_i]$

[Ans: d]

ii. You are given the actual values of these observations:

$$x_1 = 1, \quad x_2 = 2, \quad x_3 = 3, \quad x_4 = 4, \quad x_5 = 5, \quad x_6 = 6$$

What is the maximum likelihood estimate for λ? Enter your answer correct to three decimal places.

[Ans: 0.25 to 0.35]

2. For the given Beta distribution, choose the correct option. (MSQ)

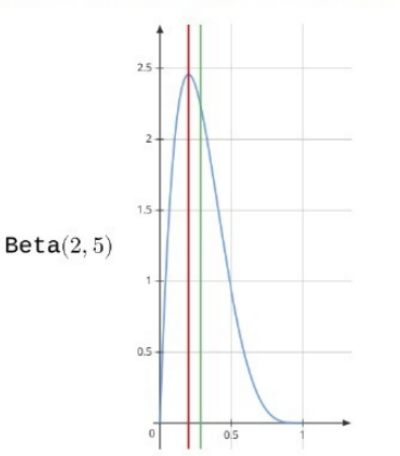


Figure 1: Beta(2,5)

(a) Red line denotes the mean for the given Beta distribution, i.e. mean $= \frac{2}{7}$

(b) Green line denotes the mean for the given Beta distribution, i.e. mean $= \frac{2}{7}$

(c) Red line denotes the mode for the given Beta distribution, i.e. mode $= \frac{1}{5}$

(d) Green line denotes the mode for the given Beta distribution, i.e. mode $= \frac{1}{5}$

[Ans: b, c]

3. Consider a dataset with 100 total data points. Each data point is classified as either type A or type B. We model this using a Bernoulli distribution, where p is the probability of a data point being type A. If the maximum likelihood estimate (MLE) of p based on the dataset is 0.4, how many data points of type B are there in this dataset?

[Ans: 60]

May 2024 Quiz 1

4. The EM algorithm is run on a dataset of 100 points modeled using a GMM with three components. Let λ_k^i be the values obtained in the E-step in the final iteration. Here, i is the index of the data-point and k is the index of a component. Also, let $\pi_1 = 0.3, \pi_2 = 0.5, \pi_3 = 0.2$ be the estimates of the mixture probabilities. x_p represent one of the data-points, for some constant p. If $\lambda_1^p + \lambda_3^p = 0.2$, what is the probability that the second component generated x_p given that we have seen the data-point?

[Ans: 0.8]

5. **Comprehension:**
In the context of Bayesian estimation, consider a Beta prior for the parameter p of a Bernoulli distribution:

$$p \sim \text{Beta}(4, 3)$$

The dataset has 8 ones and 5 zeros.
Based on the above data, answer the given subquestions.

i. What is the posterior?

   (a) Beta$(12, 8)$
   (b) Beta$(11, 7)$
   (c) Br(0.5)
   (d) Beta$(8, 5)$

[Ans: a]

ii. If we use the expected value of the posterior as a point-estimate for the parameter of the Bernoulli distribution, what is $\hat{p}$?

[Ans: 0.6]

6. **Comprehension:**
Consider a dataset of n data-points all of which are non-negative integers. These data-points are sampled from a Poisson distribution, whose probability mass function is given below:

$$f(x;\lambda) = \frac{e^{-\lambda}\lambda^{x}}{x!}, \quad \lambda > 0$$

Here λ is a parameter.
Based on the above data, answer the given subquestions.

i. Find the log-likelihood of the dataset under this distribution.

(a) $\sum_{i=1}^{n}[-\lambda + x_i \log \lambda - \log(x_i!)]$

(b) $\prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}$

(c) $\sum_{i=1}^{n}[-\lambda + \lambda \log x_i - \log(x_i!)]$

(d) $\prod_{i=1}^{n}[-\lambda + x_i \log \lambda - \log(x_i!)]$

[Ans: a]

ii. Consider a dataset that has 25 data-points. The data-point x_i and its frequency is given in the following table:

x_i	Frequency
0	1
1	4
2	6
3	9
4	5

In case the table is not clear: the value 0 appears once in the dataset, the value 1 appears four times in the dataset, and so on. Find the maximum likelihood estimate for the parameter λ of the Poisson distribution given this dataset.

[Ans: 2.52]

Sep 2024 Quiz 1

7. Let $X_1, X_2, \ldots, X_n$ be n i.i.d. samples with parameter θ, which follows one of the following PDFs:
For $\theta = -1$, we have

$$f(x \mid \theta) = \begin{cases} 5x^4, & \text{if } 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

For $\theta = 1$, we have

$$f(x \mid \theta) = \begin{cases} 1, & \text{if } 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

Suppose we wish to find the maximum likelihood estimate of θ, then which among the following are true? (MSQ)

(a) If $\prod_{i=1}^{n} 5x_i^4 < 1$, then $\hat{\theta}_{MLE} = 1$

(b) If $\prod_{i=1}^{n} 5x_i^4 > 1$, then $\hat{\theta}_{MLE} = 1$

(c) If $\prod_{i=1}^{n} 5x_i^4 < 1$, then $\hat{\theta}_{MLE} = -1$

(d) If $\prod_{i=1}^{n} 5x_i^4 > 1$, then $\hat{\theta}_{MLE} = -1$

[Ans: a, d]

8. Consider a GMM for 5 points:

$$x_1 = 1, \quad x_2 = 1.2, \quad x_3 = 2, \quad x_4 = 1.5, \quad x_5 = 0.5$$

At some time-step in the EM algorithm, following are the values of λ_k^i for the k-th mixture after the E-step:

$$\lambda_k^1 = 0.3, \quad \lambda_k^2 = 0.1, \quad \lambda_k^3 = 2.5, \quad \lambda_k^4 = 0.6, \quad \lambda_k^5 = 0.8$$

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

[Ans: 1.50 to 1.60]

9. Consider a dataset with 100 total data points. Each data point is classified as either type A or type B. We model this using a Bernoulli distribution, where p is the probability of a data point being type A. If the maximum likelihood estimate (MLE) of p based on the dataset is 0.4, how many data points of type B are there in this dataset?

[Ans: 60]

Jan 2025 Quiz 1

10. Suppose you toss a coin four times and observe no heads. You then give the coin to your friend, who tosses it until the first head occurs. Your friend tosses the coin a total of five times. Let p denote the probability that the coin comes up heads. Find the maximum likelihood estimate of p.

[Ans: 0.10 to 0.12]

11. Suppose we fit a Gaussian Mixture Model with $K = 2$ for the following dataset:

$$D = \{x_1 = 1, x_2 = 3, x_3 = 2, x_4 = 5\}.$$

At the beginning of the t-th time step of the EM algorithm, we have $\theta^{(t)}$ as follows:

$$\begin{aligned} \pi_1 &= 0.5, \quad \pi_2 = 0.5 \\ \mu_1 &= 1, \quad \mu_2 = 3 \\ \sigma_1^2 &= 1, \quad \sigma_2^2 = 1 \end{aligned}$$

What is the probability that x_1 belongs to the first component? Enter the answer correct to two decimal places.

[Ans: 0.86 to 0.90]

12. **Comprehension:**
Suppose that 100 items are sampled from a manufacturing process and 10 are found to be defective. Let p denote the proportion of defective items. Assume that the prior distribution for p is Beta$(1, 1)$.
Based on the above data, answer the given subquestions.

i. Select the correct options from the following:

   (a) Posterior distribution for p is Beta$(91, 11)$.
   (b) Posterior distribution for p is Beta$(11, 91)$.
   (c) Posterior distribution for p is Bernoulli(0.1).
   (d) Posterior distribution for p is Bernoulli(0.9).

[Ans: b]

ii. Find the posterior mean. Enter the answer correct to two decimal places.

[Ans: 0.10 to 0.11]

May 2025 Quiz 1

13. You are presented with a dataset that has latent variables that influences your data. You are asked to use Expectation-Maximization algorithm to best capture the data. How would you define the steps in the EM (Expectation and Maximization) algorithm?

(a) E-step: Estimate the latent variables in the dataset, M-step: Maximize the likelihood over the model parameters.

(b) E-step: Estimate the number of latent variables in the dataset, M-step: Maximize the likelihood over the model parameters.

(c) E-step: Estimate the likelihood over the model parameters, M-step: Maximize the number of latent variables in the dataset.

(d) E-step: Estimate the likelihood over the model parameters, M-step: Maximize the number of parameters in the model.

[Ans: a]

14. Which of the following statements is/are true? (MSQ)

(a) EM algorithm converges to a global optimum.

(b) Lloyd's algorithm converges to a local optimum.

(c) Poor initialization does not affect the clusters produced by EM algorithm.

(d) Poor initialization leads to a slower convergence of Lloyd's algorithm.

(e) Lloyd's algorithm is sensitive to the number of clusters (k).

[Ans: b, d, e]

15. Let $X_1, X_2, \ldots, X_n$ be an i.i.d. sample drawn from a Pareto distribution with the following probability density function:

$$f(x \mid \theta) = \theta x^{-\theta-1}, \quad x \geq 1, \quad \theta > 1.$$

You are given the following sample of 5 observations:

$$X = \{2.5, 1.8, 3.2, 2.0, 2.7\}.$$

What is the maximum likelihood estimator of θ for the above sample? Enter the answer correct to three decimal places.

[Ans: 1.147 to 1.150]

2024 Apr 28 ET

16. Consider a dataset of n observations $\{x_1, x_2, \ldots, x_n\}$, where each x_i follows a Bernoulli distribution with parameter p, i.e., $x_i \sim \text{Bernoulli}(p)$ for $i = 1, 2, \ldots, n$. However, you have reason to believe that the parameter p might differ for two distinct groups within the dataset. You suspect that there are two groups in the dataset, each with its own parameter (p_1 and p_2). Now, develop an algorithm to estimate the parameters p_1 and p_2 using maximum likelihood estimation. Then, apply your algorithm to a dataset with the following observations and corresponding group labels: $\{0, 1, 1, 0, 1\}$ and $\{1, 0, 1, 0, 1\}$ for group 1 and group 2 respectively. Calculate the maximum likelihood estimates of p_1 and rounded to two decimal places.

[Ans: 0.6]

2024 Sep 01 ET

17. Consider a dataset for an unsupervised learning problem in which each data-point is either 1 or 0. This dataset is modeled using a Bernoulli distribution with parameter p. The MLE for p is 0.25. If the number of ones in the dataset is 10, find the number of zeros.

[Ans: 30]

2024 Dec 22 ET

18. A train running between two stations A and B will be late on any day by a random amount X, where $X \sim \text{Uniform}[0, \theta]$. Suppose the train is late by random amounts (in minutes)

$$10, 55, 17, 20, 49, 51, 15, 4, 32, 35$$

independently on 10 days. Consider a Uniform$[0, 60]$ prior for the parameter θ. Find the maximum aposteriori (MAP) estimate of θ.
Recall that $\hat{\theta}_{MAP} = \arg\max_\theta f(\theta \mid \{X_1, \ldots, X_n\})$, where f is the posterior distribution.

[Ans: 55]

2025 Apr 13 ET

19. A coin is flipped 1000 times, and it lands on heads 150 times. You use a Beta$(10, 40)$ prior to estimate the true probability of heads. Given these observations, calculate the posterior mean of the probability of heads. Enter the answer correct to three decimal places.
Note:

- Pdf of Beta distribution is $f(\theta; \alpha, \beta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)}$, $0 \leq \theta \leq 1, \alpha, \beta \geq 0$, where $B(\alpha, \beta)$ is the Beta function.
- The mean of the Beta distribution is: $E[\theta] = \frac{\alpha}{\alpha+\beta}$.

[Ans: 0.148 to 0.154]

2025 Aug 31 ET

20. A machine's lifetime (in hours) is modeled as a random variable $X \sim$ Exponential(λ). Suppose the observed lifetimes (in hours) of the machine over 8 trials are:

$$5.2, \quad 0.8, \quad 3.4, \quad 2.5, \quad 1.1, \quad 4.0, \quad 6.3, \quad 2.9$$

Assume that the parameter λ has a prior distribution given by $\lambda \sim$ Uniform$(0, 2)$. Find the maximum a posteriori (MAP) estimate of λ. Enter the answer correct to two decimal places.
Recall that:

$$\hat{\lambda}_{\text{MAP}} = \arg\max_{\lambda} f(\lambda \mid \{X_1, \ldots, X_n\}),$$

where f is the posterior distribution.

[Ans: 0.28 to 0.34]

2025 Dec 18 ET

21. Suppose we have a Uniform$(1, 10)$ prior for the parameter θ. Then, the value of the maximum likelihood estimate (MLE) is equal to the value of the maximum a posteriori (MAP) estimate.

(a) True

(b) False

[Ans: a]

2026 May 06 ET

22. A coin is flipped 100 times, and it lands on heads 60 times. You use a Beta$(10, 40)$ prior to estimate the true probability of heads. Given these observations, calculate the posterior mean of the probability of heads. Enter the answer correct to three decimal places.
Note:

- Pdf of Beta distribution is $f(\theta; \alpha, \beta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)}$, $0 \leq \theta \leq 1, \alpha, \beta \geq 0$, where $B(\alpha, \beta)$ is the Beta function.
- The mean of the Beta distribution is: $E[\theta] = \frac{\alpha}{\alpha+\beta}$.

[Ans: 0.44 to 0.49]