ECE657 Tools of Intelligent Systems Design
Assignment 4 Question 1

Huy Trinh

July 2024

Question 1: MLE and MAP Estimation

Q1.1

Liklihood functiuon $L(\lambda)$ is defined as

$$L(\lambda) = \prod_{i=1}^{N} p(x^i) = \prod_{i=1}^{N} \mathrm{e}^{-\lambda x}.$$

The probability density function of the exponential distribution is given by:

$$p(x; \lambda) = \lambda e^{-\lambda x} \quad \text{for } x \geq 0$$

For $n = 5$, given the dataset $X = \{2.1, 2.5, 1.8, 2.4, 2.0\}$ and substitute the formula above, the likelihood function is:

$$L(\lambda) = \lambda e^{-\lambda \cdot 2.1} \cdot \lambda e^{-\lambda \cdot 2.5} \cdot \lambda e^{-\lambda \cdot 1.8} \cdot \lambda e^{-\lambda \cdot 2.4} \cdot \lambda e^{-\lambda \cdot 2.0}$$

$$L(\lambda) = \lambda^5 e^{-\lambda \sum_{i=1}^{5} x_i}$$

2. **Derive the log-likelihood function $\ell(\lambda)$.**

The log-likelihood function $\ell(\lambda)$ is the natural logarithm of the likelihood function:

$$\ell(\lambda) = \log L(\lambda) = \log \left(\lambda^5 e^{-\lambda \sum_{i=1}^{5} x_i} \right)$$

$$\ell(\lambda) = \log(\lambda^5) + \log \left(e^{-\lambda \sum_{i=1}^{5} x_i} \right)$$

$$\ell(\lambda) = 5 \log \lambda + \log \left(e^{-\lambda \sum_{i=1}^{5} x_i} \right)$$

$$\ell(\lambda) = 5 \log \lambda - \lambda \sum_{i=1}^{5} x_i$$

3. **Determine the maximum likelihood estimate (MLE) for the rate parameter λ.**

Finding MLE: we take the derivative of the log-likelihood function with respect to λ and set it to zero:

$$\frac{d}{d\lambda}\ell(\lambda) = \frac{d}{d\lambda}\left(5\log\lambda - \lambda\sum_{i=1}^{5} x_i\right) = 0$$

$$\leftrightarrow \frac{5}{\lambda} - \sum_{i=1}^{5} x_i = 0$$

$$\leftrightarrow \frac{5}{\lambda} = \sum_{i=1}^{5} x_i$$

$$\leftrightarrow \lambda = \frac{5}{\sum_{i=1}^{5} x_i}$$

4. **Calculate the MLE using the provided data.**
The provided data is $X = \{2.1, 2.5, 1.8, 2.4, 2.0\}$:

$$\sum_{i=1}^{5} x_i = 2.1 + 2.5 + 1.8 + 2.4 + 2.0 = 10.8$$

$$\lambda = \frac{5}{10.8} \approx 0.463$$

Q1.2

1. **Write down the prior distribution $P(\lambda)$ for the Gamma distribution with a shape parameter $\alpha = 2$ and a rate parameter $\beta = 1$.**

The probability density function of the Gamma distribution is given by:

$$P(\lambda) = \frac{\beta^\alpha \lambda^{\alpha-1} e^{-\beta\lambda}}{\Gamma(\alpha)}$$

For $\alpha = 2$ and $\beta = 1$:

$$P(\lambda) = \frac{1^2 \lambda^{2-1} e^{-1\lambda}}{\Gamma(2)} = \frac{\lambda e^{-\lambda}}{1} = \lambda e^{-\lambda}$$

2. **Derive the log-posterior distribution $\log P(\lambda|X)$ by combining the log-likelihood function from the previous sub-question with the log-prior.**

The posterior distribution is found by applying Bayes' theorem, which involves multiplying the likelihood by the prior probability and normalizing:

$$P(\lambda|X) \propto L(\lambda)P(\lambda)$$

The log-posterior is:

$$\log P(\lambda|X) = \ell(\lambda) + \log P(\lambda)$$

Substitute the log-likelihood function above, we get:

$$\log P(\lambda|X) = (5\log\lambda - \lambda\sum_{i=1}^{5} x_i) + (\log\lambda - \lambda)$$

$$\log P(\lambda|X) = 6\log\lambda - \lambda\left(\sum_{i=1}^{5} x_i + 1\right)$$

$$\log P(\lambda|X) = 6\log\lambda - 11.8\lambda$$

3. **Determine the maximum a posteriori (MAP) estimate for the rate parameter λ.**

Finding MAP estimate: take the derivative of the log-posterior with respect to λ and set it to zero:

$$\frac{d}{d\lambda}\log P(\lambda|X) = \frac{d}{d\lambda}(6\log\lambda - 11.8\lambda) = 0$$

$$\leftrightarrow \frac{6}{\lambda} - 11.8 = 0$$

$$\leftrightarrow \lambda = \frac{6}{11.8} \approx 0.508$$

Solving for λ:

$$\frac{5}{\lambda} = \sum_{i=1}^{5} x_i$$

$$\lambda = \frac{5}{\sum_{i=1}^{5} x_i}$$

4. **Calculate the MAP estimate using the provided data.**

The MAP estimate using the provided data is:

$$\lambda \approx 0.508$$

Q1.3:

Some main differences

* **Maximum Likelihood Estimation (MLE):** By definition, the maximum likelihood estimation is the parameter that maximize the Likelihood function which represents the probability of observing the given data under different parameter values. It only emphasizes on **observed data** without taking prior data into consideration.

$$\hat{\lambda}_{\text{MLE}} = \arg\max_{\lambda} L(\lambda|X)$$

When data is scarce, MLE estimates can be unstable and highly sensitive to the observed data. This is due to MLE does not have any additional information to rely on beyond the observed data (low varaince)

* **Maximum A Posterior (MAP):** By definition, the Maximum A Posterior is parameter that maximize the posterior distribution, which combines the likelihood of the data and the prior distribution of the parameter. It take into account the prior data in addition to observed data for prior knowledge, which means that the likelihood is weighted by the prior in MAP.

$$\hat{\lambda}_{\text{MAP}} = \arg\max_{\lambda} P(\lambda|X) = \arg\max_{\lambda} (L(\lambda|X)P(\lambda))$$

When data is scarce, MAP can be more stable due to its posterior distribution combines both the observed data and prior information. The prior can act as a regularizer, guiding the estimation towards more plausible values based on prior knowledge.