

Short Answer and “True or False” Conceptual Questions

Maximum Likelihood Estimator(MLE)

(A.1)

(A.1.a)

Objective: Find the expression for the maximum=likelihood estimate for the parameter λ for the poisson distribution, interns of the goal count. Assume idd rvs.

Here we will assume that observations obtained takes the form x_1, x_2, \dots, x_N , and then we derive the best estimator for λ in this much general context.

I will shut up and just show you the math:

$$\begin{aligned} & \prod_{n=1}^N \text{Poi}(x_n|\lambda) \\ & \sum_{n=1}^N \log(\text{Poi}(x_n|\lambda)) \\ & \sum_{n=1}^N (-\lambda + x_n \ln(\lambda) + \log(x_n!)) \end{aligned} \tag{A.1.a.1}$$

Notice that, only some of the terms are relevant to the parameter λ , therefore, the optimization problem we are solving is:

$$\lambda^+ = \underset{\lambda}{\operatorname{argmax}} \left\{ -N\lambda + \ln(\lambda) \sum_{i=1}^N (x_n) \right\} \tag{A.1.a.2}$$

To solve it, we just take the derivative, set it to zero and then solve for λ , because this function is a function that has a single local maximum.

$$\begin{aligned} \partial_{\lambda} \left[-N\lambda + \ln(\lambda) \sum_{i=1}^N (x_n) \right] &= 0 \\ -N + \frac{\sum_{n=1}^N x_n}{\lambda^+} &= 0 \\ \implies \lambda^+ &= \frac{\sum_{n=1}^N x_n}{N} \end{aligned} \tag{A.1.a.3}$$

Therefor, for this particular problem, the best estimator will be the average of all the observation, which is just:

$$\frac{2 + 4 + 6 + 1}{5} = 2.6$$

And that is the answer for the question.

(A.1.b)

The derivation of the best estimator in the general context is shown in [A.1.a](#).

The numerical value for six observations is:

$$\frac{2 + 4 + 6 + 1 + 3}{7} = 2.666666666 \dots$$

(A.1.c)

The numerical results for 5 observations has been shown in [A.1.a](#) and [A.1.b](#) respectively.

(A.2)

Objective: Find the MLE for the uniform distribution on $[0, \theta]$, where θ is the value we want to estimate.

Suppose that an observations has been made: x_1, x_2, \dots, x_N , and we assume that they are iid, and we want to find the likelihood of such an observation is generated using the Uniform distribution. And this will be given by the expression:

$$\prod_{n=1}^N \underbrace{\mathbb{P}(X = x_n)}_{\frac{1}{\theta} \mathbf{1}_{\{0 \leq x_n \leq \theta\}}} \quad (\text{A.2.1})$$

Observe that, if any of the observation is beyond the range $[0, \theta]$, we will have zero likelihood, so let's assume that $\theta \leq \max_{1 \leq i \leq N}(x_i)$, then we will have this expression for the likelihood:

$$\prod_{n=1}^N \frac{1}{\theta} = \frac{1}{\theta^N} \quad (\text{A.2.2})$$

And taking the log on that we have:

$$\log\left(\frac{1}{\theta^N}\right) = -N \log(\theta) \quad (\text{A.2.3})$$

Observe that that function is monotonically decreasing as the value of θ get larger and larger, therefore, to maximize the likelihood, we need the value of θ to be as small as possible, and the smallest possible such θ that is not giving us zero likelihood is: $\max_{1 \leq i \leq N}(x_i)$, therefore, best estimate is given by:

$$\theta^+ = \max_{1 \leq i \leq N}(x_i)$$

Overfitting

Bias-Variance Tradeoff

Polynomial Regression

Ridge Regression on MNIST