Ass4 Task 1

30 June 2023 11:22

1 Maximum Likelihood Estimation [5 points]

We say that $X \in \{0, 1, ..., \}$ has a Poisson distribution with rate parameter $\lambda > 0$, written $X \sim \mathrm{Poi}(\lambda)$, if its probability mass function is

$$\mathrm{Poi}(\lambda) = \frac{\lambda^x}{x!} e^{-\lambda}$$

The term $e^{-\lambda}$ is just a normalization constant required to ensure the distribution sums to 1. The Poisson distribution is often used as a model for counts of rare events like radioactive decay and accidents.

Tasks:

1. Let us assume a dataset with i.i.d. observations $X = \{x_1, \dots, x_N\}$ and a Poisson distribution. Derive the maximum likelihood estimate for the rate parameter λ using

$$p(X|\lambda) = \prod_{i=1}^{N} \frac{\lambda^{x_i}}{x_i!} e^{-\lambda}$$

by optimizing the log-likelihood (compute the derivative, set it to zero, and solve for λ).

2. Why do we transform the likelihood into log-likelihood in the process?

1) Step 1: Compute derivative of $p(X|\lambda)$

Step 1.1: Transform into Log-likelihood

$$p(X|\lambda) = \prod_{i=1}^{N} \frac{\lambda^{x_i}}{x_i!} \cdot e^{-\lambda}$$

$$\ln(p(X|\lambda)) = \ln\left(\prod_{i=1}^{N} \frac{\lambda^{x_i}}{x_i!} \cdot e^{-\lambda}\right)$$

$$= \sum_{i=1}^{N} \Big(\underbrace{\ln\left(\frac{\lambda^{x_i}}{x_i!}\right)}_{\substack{= \ln(\lambda^{x_i}) - \ln(x_i!) \\ = x_i \cdot \ln(\lambda) - \ln(x_i!)}} + \underbrace{\ln(e^{-\lambda})}_{= -\lambda} \Big)$$

$$L(\lambda) = \sum_{i=1}^{N} (x_i \cdot \ln(\lambda) - \ln(x_i!) - \lambda)$$

Step 1.2 Calculate derivative

$$\frac{\partial L(\lambda)}{\partial \lambda} = \frac{\partial}{\partial \lambda}\left(\sum_{i=1}^{N} (x_i \cdot \ln(\lambda) - \ln(x_i!) - \lambda)\right)$$

$$= \sum_{i=1}^{N} \Big(\underbrace{\frac{\partial}{\partial \lambda}(x_i \cdot \ln(\lambda))}_{= \frac{x_i}{\lambda}} - \underbrace{\frac{\partial}{\partial \lambda}(\ln(x_i!))}_{= 0} - \underbrace{\frac{\partial}{\partial \lambda}(\lambda)}_{= 1} \Big)$$

$$= \sum_{i=1}^{N} \left(\frac{x_i}{\lambda} - 1\right)$$

Step 2: Set derivative to 0 and solve for λ

$$\sum_{i=1}^{N} \left(\frac{x_i}{\lambda} - 1\right) = 0 \quad | +1$$

$$\sum_{i=1}^{N} \left(\frac{x_i}{\lambda}\right) = \sum_{i=1}^{N} 1 = N \quad | : N$$

$$\sum_{i=1}^{N} \left(\frac{x_i}{N\lambda}\right) = 1 \quad | \cdot \lambda$$

$$\lambda = \sum_{i=1}^{N} \left(\frac{x_i}{N}\right)$$

2) Why do we transform the likelihood into log-likelihood?

-> Logarithm is a monotonic transformation that preserves the locations of the extrema (in particular, the estimated parameters in max-likelihood are identical for the original and the log-transformed formulation)

Simplicity: In some problems it is easier to work with the log-likelihood function. Taking the logarithm of the likelihood function simplifies the mathematical expression by converting products (from the likelihood) into sums (in the log-likelihood). This simplification makes it easier to differentiate the function.
Numerical stability:
The computer uses a limited digit floating point representation of fractions, multiplying many probabilities that might be very small can lead to an underflow. With log and getting rid of the multiplications the computer doesn't have this issue.