Parametric distributions

Part 1: Pen and paper exercise

Verify the ML estimator for the Bernoulli distribution.

Setup: Given the Bernoulli distribution

$$\text{Bern}(x|\mu) = \mu^x (1-\mu)^{1-x} \tag{1}$$

and independent samples $D = \{x_1, \ldots, x_n\}$ with $x_i \in \{0,1\}$ from the distribution, we want to estimate the parameter μ of the distribution.

The likelihood function is:

$$p(D|\mu) = \prod_{n=1}^{N} p(x_n|\mu) = \prod_{n=1}^{N} \text{Bern}(x_n|\mu) = \prod_{n=1}^{N} \mu^{x_n} (1-\mu)^{1-x_n} \tag{2}$$

We want to find the parameter μ that maximizes the likelihood function. Since the logarithm does not change the maximum, we can maximize the logarithm of the likelihood function instead.

$$\mu^* = \text{argmax}_\mu \ln p(D|\mu) \tag{3}$$

$$\ln p(D|\mu) = \sum_{n=1}^{N} \Big(x_n \ln \mu + (1-x_n) \ln(1-\mu) \Big) \tag{4}$$

Solution: To find the maximum of a function, we find the only zero point of the first derivative of the function and show that the second derivative is negative at that point. Since our optimization is bounded with $\mu \in [0,1]$, this guarantees a global maximum.

To find the derivative see product rule, quotient rule, chain rule and derivative of the logarithm function as $\frac{\partial \ln(x)}{\partial(x)} = \frac{1}{x}$.

We define $m = \sum_{n=1}^{N} x_n$ as the number of “heads” or positive outcomes.

1) Find the derivative of the likelihood function.

$$\begin{aligned}
\frac{\partial \ln p(D|\mu)}{\partial \mu} &= \sum_{n=1}^{N} \left(\frac{x_n}{\mu} - \frac{1-x_n}{1-\mu} \right) && (5) \\
&= \frac{1}{\mu} \sum_{n=1}^{N} x_n - \frac{1}{1-\mu} (N - \sum_{n=1}^{N} x_n) && (6) \\
&= \frac{m}{\mu} - \frac{N-m}{1-\mu} && (7)
\end{aligned}$$

2) Solve for zero to find μ^*.

$$\frac{m}{\mu^*} - \frac{N-m}{1-\mu^*} = 0 \tag{8}$$
$$\frac{m}{\mu^*} = \frac{N-m}{1-\mu^*} \tag{9}$$
$$m(1-\mu^*) = (N-m)\mu^* \tag{10}$$
$$m - m\mu^* = N\mu^* - m\mu^* \tag{11}$$
$$m = N\mu^* \tag{12}$$
$$\frac{m}{N} = \mu^* \tag{13}$$

3) Now we have an extreme point (minimum, maximum or saddlepoint). To show that we have indeed found the maximum we will show that the second derivative is negative at the point μ^*. First, find the second derivative of the likelihood.

$$\frac{\partial^2 \ln p(D|\mu)}{\partial \mu^2} = \frac{\partial}{\partial \mu}\Big(\frac{m}{\mu} - \frac{N-m}{1-\mu}\Big) \tag{14}$$
$$= -\frac{m}{\mu^2} - \frac{N-m}{(1-\mu)^2} \tag{15}$$
$$\tag{16}$$

4) This is already negative for all real values of μ since both fractions will always be positive. However, we can substitute our point μ^*:

$$-\frac{m}{(\mu^*)^2} - \frac{N-m}{(1-\mu^*)^2} = -\frac{m}{(\frac{m}{N})^2} - \frac{N-m}{(1-\frac{m}{N})^2} \tag{17}$$
$$= -\frac{N^2}{m} - \frac{N(1-\frac{m}{N})}{(1-\frac{m}{N})^2} \tag{18}$$
$$= -\frac{N^2}{m} - \frac{N}{(1-\frac{m}{N})} \tag{19}$$
$$= -\frac{N^2}{m} - \frac{N^2}{(N-m)} \tag{20}$$

This is clearly negative since $0 \leq m \leq N$.

For the edge cases where $m = 0$ or $m = N$, the second derivative will approach $-\infty$, which is also negative.

Therefore we have found the maximum of the likelihood function.