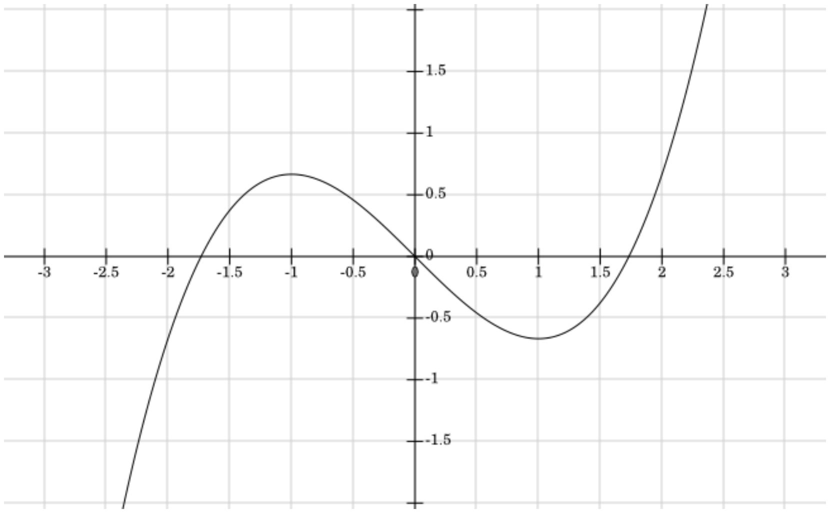
Maximum likelihood estimation exercise

Maxima and minima



$$f(x) = \frac{1}{3}x^3 - x$$

$$\frac{\partial}{\partial x} f(x) = x^2 - 1 = 0 \quad \Rightarrow \quad x = \pm 1$$ Two stationary points

$$\frac{\partial^2}{\partial x^2} f(x) = \frac{\partial}{\partial x}\left(\frac{\partial}{\partial x} f(x)\right) = \frac{\partial}{\partial x}\left(x^2 - 1\right) = 2x \quad \Rightarrow$$

$2x > 0, \text{ for } x = 1$ Local minimum

$2x < 0, \text{ for } x = -1$ Local maximum

(No global maxima or minima)

Maximum likelihood estimation

Example: consider an experiment consisting on tossing a coin. The outcomes of *n* tosses are:

$$x_1, \ldots, x_n$$

Where $x_i = 1$ for heads, and $x_i = 0$ for tails.

This is a single coin tossing model with parameter p , i.e. probability of obtaining heads. The likelihood after n tosses is:

$$L = P(x_1) \cdots P(x_n) = p^{x_1}(1-p)^{(1-x_1)} \cdot \ldots \cdot p^{x_n}(1-p)^{(1-x_n)} = p^{\sum_{i=1}^{n} x_i}(1-p)^{\sum_{i=1}^{n}(1-x_i)} = p^h(1-p)^{n-h}$$

The extreme value of a function is the solution to the condition of zero derivative

$$\frac{\partial}{\partial p} L = hp^{h-1}(1-p)^{n-h} - (n-h)(1-p)^{n-h-1} p^h = 0$$

Maximum likelihood estimation

Example: consider an experiment consisting on tossing a coin. The outcomes of *n* tosses are:

$$x_1, \ldots, x_n$$

Where $x_i = 1$ for heads, and $x_i = 0$ for tails.

This is a single coin tossing model with parameter p , i.e. probability of obtaining heads. The likelihood after n tosses is:

$$L = P(x_1) \cdots P(x_n) = p^{x_1}(1-p)^{(1-x_1)} \cdot \ldots \cdot p^{x_n}(1-p)^{(1-x_n)} = p^{\sum_{i=1}^{n} x_i}(1-p)^{\sum_{i=1}^{n}(1-x_i)} = p^h(1-p)^{n-h}$$

The extreme value of a function is the solution to the condition of zero derivative

$$\frac{\partial}{\partial p} L = hp^{h-1}(1-p)^{n-h} - (n-h)(1-p)^{n-h-1} p^h = 0$$

Dividing by $p^h(1-p)^{n-h}$

$$hp^{-1} - (n-h)(1-p)^{-1} = 0$$

Maximum likelihood estimation

Example: consider an experiment consisting on tossing a coin. The outcomes of *n* tosses are:

$$x_1, \ldots, x_n$$

Where $x_i = 1$ for heads, and $x_i = 0$ for tails.

This is a single coin tossing model with parameter p , i.e. probability of obtaining heads. The likelihood after n tosses is:

$$L = P(x_1)\cdots P(x_n) = p^{x_1}(1-p)^{(1-x_1)} \cdot \ldots \cdot p^{x_n}(1-p)^{(1-x_n)} = p^{\sum_{i=1}^{n} x_i}(1-p)^{\sum_{i=1}^{n}(1-x_i)} = p^h (1-p)^{n-h}$$

The extreme value of a function is the solution to the condition of zero derivative

$$\frac{\partial}{\partial p} L = hp^{h-1}(1-p)^{n-h} - (n-h)(1-p)^{n-h-1} p^h = 0$$

Dividing by $p^h(1-p)^{n-h}$

$$hp^{-1} - (n-h)(1-p)^{-1} = 0$$

$$h(1-p) = (n-h)p \quad \longrightarrow \quad p = \frac{h}{n}$$

We leave as exercise to proof that this is indeed a maximum

Maximum likelihood estimation

The probability estimation that maximizes the likelihood is the one obtained from calculating the frequencies of events.

In general it is easier to calculate the maxima for the log-likelihood:

$$\log L \qquad\qquad \frac{\partial}{\partial p}\log L = \frac{1}{L}\frac{\partial L}{\partial p}$$

The log function is a monotonically increasing function; hence, log L it attains the maxima at the same values as L.

In more complex situations, optimization techniques are applied, like gradient descent, Montecarlo with simulated annealing, expectation maximization (EM), etc..

The likelihood is not a probability distribution. So we should not say “the likelihood of the data”. Always say “the likelihood of the parameters”