7.5 Parameter ESTIMATION: MLE

MLE: Maximum Likelihood estimate

Setup:

Data: Given a data set D containing n data points $X_1, X_2, ..., X_n$ denoted by $D = \{X_1, X_2, ..., X_n\}$. Each data points X_i is d dimensional, i.e., X_i has d attributes $x_{i1}, x_{i2}, ..., x_{id}$ and each attribute is a real number, denoted by $X_i = (x_{i1}, x_{i2}, ..., x_{id}) \in \mathbb{R}^d$. The whole data set can be visualized as a table having ***d*** columns in a row and a total of ***n*** rows.

Assume we have a distribution P, which is having parameter θ.

$$\{P_\theta | \theta \in \{mean, variance, \ldots.\}\},$$

P_θ is a distribution having a set of parameter denoted by θ.

Assume that the data D: $\{x_1, x_2, ..., x_n\}$ is drawn from this distribution in independent and identical manner, i.e., i.i.d. We denoted it as $D \underset{i.i.d}{\sim} P_\theta$, for some parameter θ.

Goal: is to estimate the value of θ that D comes from. The estimated value of θ is called as θ_{MLE} and it has following property.

$$\theta_{MLE} = \underset{\theta}{argmax}\, P(D|\theta)$$

It says that θ_{MLE} is that value, of parameter, which maximizes the probability of data "D" given θ.

$$\begin{aligned} D &= \{X_1, X_2, \ldots, X_n\}, \\ P\{D|\theta\} &= P(X_1, X_2, \ldots, X_n|\theta) \\ &= P(X_1|\theta)\,P(X_2|\theta)\ldots P(X_n|\theta) \\ &= \prod_{i=1}^{n} P(X_i|\theta). \end{aligned}$$

All X_i are conditionally independent given θ.

To get the maximum likelihood estimator of θ, θ_{MLE}, we first take the logarithm of $\prod_{i=1}^{n} P(X_i|\theta)$, then:

$$\log \prod_{i=1}^{n} P(X_i|\theta) = \sum_{i=1}^{n} P(X_i|\theta)\ ,$$

then we take partial differentiate with respect to θ and equate it to 0.

$$\frac{\partial \log \prod_{i=1}^{n} P(X_i|\theta)}{\partial\theta} = \frac{\partial \sum_{i=1}^{n} P(X_i|\theta)}{\partial\theta} = 0\,,$$

solve it to get the θ_{MLE}.

Note : P(D|θ) is called likelihood of the data given θ.

Having seen the mathematical background of MLE let’s see what are the MLE’s for some of the popular distributions.

Normal Distribution: Probability density function for univariate normal distribution is given by the following equation. Univariate means that data is one dimensional, i.e., data is having only one attribute, e.g., “rv”.

$$f(x|\theta : (\mu, \sigma^2)) = \frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$
$$Where,$$
$$\mu\, is\, mean\, or\, expectation\, of\, distribution$$
$$\sigma\, is\, standard\, deviation$$
$$\sigma^2\, is\, variance$$
$$\mu_{MLE} = \frac{1}{n}\sum x_i$$
$$where,$$
$$\frac{1}{n}\sum x_i\, is\, the\, sample\, mean\, of\, the\, data$$
$$So,\ MLE\, of\, mean\, of\, the\, normal\, distribution\, is\, it's\, sample\, mean.$$

$$\sigma^2_{MLE} = \frac{1}{n}\sum_{i=i}^{i=n}(x_i - \mu)^2, which\, is\, sample\, variance$$

Note : In above notation x is a data point.

Proof of MLE is beyond the scope of this course.

Poisson Distribution: is given by the following equation

$$P(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$
$$where,$$
$$\lambda\, is\, the\, average\, number\, of\, events\, per\, interval$$
$$e\, is\, the\, number\, 2.71828...\, the\, base\, of\, the\, natural\, logs$$
$$x\, is\, the\, number\, of\, times\, an\, event\, has\, occurred\, in\, the\, interva$$
$$\lambda_{MLE} = \frac{1}{n}\sum x_i$$
$$is\, the\, sample\, mean\, of\, the\, data$$

Similarly, for **binomial distributing** MLE of its parameter is the mean of the success**.**

Now we know the formula of MLE for various distribution it is very easy to do parameter estimation using NumPy.

Normal Distribution

```
import numpy as np
# generate 1000 random variable
rv = np.random.normal(0, 2.5, 1000)  # first parameter is mean=0, second is SD=2.5 and third is sample size=1000
# rv is a numpy ndarray

print "MLE of mean", rv.mean()
print "MLE of SD", rv.std()
```

Perform above experiment by changing size, you will notice as size increases MLE gets more and more closer to the actual mean and SD.

Poisson Distribution

```
import numpy as np
# generate 1000 random variable
poissonDistribution = np.random.poisson(2.6, 10000)  # first parameter is lambda or average number of event per interval, second is sample size=1000
# poissonDistribution is a numpy ndarray

print "MLE of lambda", poissonDistribution.mean()
```

There are some problems with MLE like:

1. It over fits the data, meaning that the outcome of the parameter estimate is sensitive to random variations in the data.
2. MLE estimation does not incorporate any prior knowledge.
3. It might not be unique.
4. There might not be an MLE at all.