Maximum Likelihood Estimation for univariate Gaussian

Sonam Kumar
sonam[dot]ibab[at]gmail[dot]com

Data: $\{x_i\}_{i=1}^{N}$ Model: $p(x_i \mid \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x_i-\mu)^2}{2\sigma^2}\right\}$

Here, we need to maximize the "loss function" (I'd like to call the same as a GAIN FUNCTION), namely the likelihood function. The function expresses the probability that the given data $\{x_i\}_{i=1}^{N}$ fits the Gaussian distribution $N(\mu, \sigma)$ for model parameters μ, σ denoting respectively the mean and the standard deviation of the distribution from which the data might have "been sampled".

Assume that the data points are independent and identically distributed random variables:

$$X = x_1, X = x_2, \ldots, X = x_{N-1}, X = x_N$$

Define the maximum likelihood estimator (MLE) function as:

$$f(\theta) = P(X \mid \theta)$$

where, $X = (x_1, x_2, \ldots, x_{N-1}, x_N)^t$ is the vector of features, and

$\theta = (\mu, \sigma)^t$ is the parameter vector, or

$\theta = \mu$, or $\theta = \sigma$ [depends on situation]

$$\text{or,} \quad f(\theta) = p(x_1, x_2, \ldots, x_{N-1}, x_N \mid \theta)$$

$$\text{or,} \quad f(\theta) = \prod_{i=1}^{N} p(x_i \mid \theta)$$

Since the random variables are independent

$$\text{or,} \quad \ln[f(\theta)] = \ln\left[\prod_{i=1}^{N} p(x_i \mid \theta)\right]$$

$$\text{or,} \quad \ln[f(\theta)] = \sum_{i=1}^{N} \ln[p(x_i \mid \theta)]$$

Since $\ln(a \times b) = \ln(a) + \ln(b)$

<u>Notation</u>: $l(\theta) = \ln[f(\theta)]$

For each $i = 1, \dots, N$ we have the following:

$$\ln\left[p(x_i|\theta)\right] = \ln\left[\frac{1}{\sqrt{2\pi}\,\sigma}\exp\left\{-\frac{(x_i-\mu)^2}{2\sigma^2}\right\}\right]$$

$$= -\ln\left[\sqrt{2\pi}\right] - \ln\left[\sigma\right] - \frac{(x_i-\mu)^2}{2\sigma^2}$$

Then, log-likelihood function $l(\theta)$ can be written as:

$$l(\theta) = \sum_{i=1}^{N}\left\{-\ln\left[\sqrt{2\pi}\right] - \ln[\sigma] - \frac{(x_i-\mu)^2}{2\sigma^2}\right\}$$

$$= \underbrace{-N\ln\left[\sqrt{2\pi}\right]}_{\text{constant}} - \underbrace{N\ln[\sigma]}_{\text{function of } \sigma} - \underbrace{\sum_{i=1}^{N}\frac{(x_i-\mu)^2}{2\sigma^2}}_{\text{function of } \mu \text{ and } \sigma}$$

Note that we are using θ as a vector of parameters μ and σ. So, in order to find the vector θ at which $l(\theta)$ attains maximum we need to solve for the zero gradient of $l(\theta)$:

$$\nabla_\theta l(\theta) = 0,$$

which is equivalent to solving the system:

$$\frac{\partial}{\partial\mu} l(\theta) = 0 \quad , \quad \frac{\partial}{\partial\sigma} l(\theta) = 0$$

Computing partial differentiation w.r.t. μ considering σ as constant:

$$-0 - 0 - \sum_{i=1}^{N}\frac{1}{2\sigma^2}\left\{2(x_i-\mu)(-1)\right\} = 0$$

or,

$$\sum_{i=1}^{N}(x_i-\mu) = 0$$

All other terms vanish because they are constants and/or independent of 'i' and because R.H.S. is 0.

Simplifying further gives:

$$\sum_{i=1}^{N} x_i - \sum_{i=1}^{N} \mu = 0$$

or, $$N\mu = \sum_{i=1}^{N} x_i$$

or, $$\boxed{\mu = \frac{1}{N} \sum_{i=1}^{N} x_i}$$

Computing partial differentiation w.r.t. σ considering μ as constant:

$$0 - N\left(\frac{1}{\sigma}\right) - \sum_{i=1}^{N} \frac{(x_i-\mu)^2}{2} \left\{(-2)\frac{1}{\sigma^3}\right\} = 0$$

or, $$N - \frac{1}{\sigma^2} \sum_{i=1}^{N} (x_i-\mu)^2 = 0$$

All other terms vanish because they are constants and/or independent of 'i' and because R.H.S. is 0.

Simplifying further gives:

$$N = \frac{1}{\sigma^2} \sum_{i=1}^{N} (x_i-\mu)^2$$

or, $$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i-\mu)^2$$

or, $$\boxed{\sigma = \sqrt{\frac{\sum_{i=1}^{N} (x_i-\mu)^2}{N}}}$$

Final Statement:

Best Gaussian fit of the data $\{x_i\}_{i=1}^{N}$ is by $N(\bar{x}, s)$, where $\bar{x}$ is sample mean and s is sample standard deviation.

Proved !!!