Freshman Physics Laboratory (PH003)

Corrections to Parameter Estimation Chapter of the Vademecum of Data Analysis Beginners

Academic Year 2008-2009

Copyright©Virgínio de Oliveira Sannibale, 2001
Chapter 6

Errata Correction Parameter Estimation

When a function depends on a set of parameters we are faced with the problem of estimating the values of those parameters. Starting from a finite set of measurements we can make a statistical determination of the parameters set.

In the following sections we will examine two standard methods, the maximum-likelihood and least-square methods, to estimate the parameters of a PDF and of a general function of one independent variable.

6.1 The Maximum Likelihood Principle (MLP)

Let $x$ be a random variable and $f$ its PDF, which depends on a set of unknown parameters $\vec{\theta} = (\theta_1, \theta_2, ..., \theta_n)$

$$f = f(x; \vec{\theta}).$$

Given $N$ independent samples of $x$, $\vec{x} = (x_1, x_2, ..., x_N)$, the quantity

$$L(\vec{x}; \vec{\theta}) = \prod_{i=1}^{N} f(x_i; \vec{\theta})$$

is called the likelihood of $f$. $L$ is proportional to the probability to obtain the set of samples $\vec{x}$, assuming that the $N$ samples are independent.
The maximum likelihood principle (MLP) states that the best estimate of the parameters $\hat{\theta}$ is the set of values which maximizes $L(\hat{x}; \hat{\theta})$.

Considering the monotonic property of the natural logarithm function, it is quite often convenient to use the equivalent expression of the likelihood function

$$L^*(\hat{x}; \hat{\theta}) = \sum_{i=1}^{N} \log \left[ f(x_i; \hat{\theta}) \right]$$

The MLP reduces the problem of parameter estimation to that of maximizing the function $L$ (or $L^*$). Because, in general, it is not possible to always find the parameters $\hat{\theta}$ that maximize $L$ or $L^*$ analytically, numerical methods implemented in computers are often used.

6.1.1 Example: $\sigma$ and $\mu$ of a Normally Distributed Random Variable

Let’s suppose we have $N$ independent samples of a normally distributed random variable $x$, whose $\sigma$ and $\mu$ are unknown. Experimentally, this case corresponds to measuring the same physical quantity $x$ several times with the same instrument.

In this case, $L^*$ is

$$L^*(\hat{x}, \hat{\theta}) = \log \left\{ \left( \frac{1}{\sqrt{2\pi\sigma}} \right)^N \exp \left[ -\sum_{i=1}^{N} \frac{(x_i - \mu)^2}{2\sigma^2} \right] \right\},$$

$$= -N \log \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^{N} (x_i - \mu)^2 + \text{const.}$$

and we have to maximize it. The following conditions

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

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

are sufficient to determine the absolute minimum of $L^*$. 
Solving the first equation respect to $\mu$, we obtain the estimator $\hat{\mu}$ of $\mu$

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^{N} x_i .$$

Solving the second equation respect to $\sigma$, we obtain the estimator $\hat{\sigma}$ of $\sigma$

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

The estimator of the variance is biased, i.e. the expectation value of the estimator is not the parameter itself; in fact, it can be demonstrated that

$$E[\hat{\sigma}^2] = \left(1 - \frac{1}{N}\right) \sigma^2 .$$

Because of this it is preferable to use the following unbiased estimator

$$s^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \hat{\mu})^2 . \quad \Rightarrow E[s^2] = \sigma^2 .$$

What is the variance associated with the average?
To answer to this question, let’s consider the average variable,

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i ,$$

which must be a Gaussian random variable. Using the pseudo-linearity property, its variance can be computed directly, i.e.

$$V[\bar{x}] = \frac{1}{N} \sigma^2 .$$