Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v8n2p267

A Simple and Conservative Empirical Likelihood Function–Corrected

By Miller

Published: 14 October 2015

Electronic Journal of Applied Statistical Analysis
Vol. 08, Issue 02, 2015, 267-269
DOI: 10.1285/i20705948v8n2p267

A Simple and Conservative Empirical Likelihood Function–Corrected

Guthrie Miller*

Santa Fe, New Mexico, USA

Published: 14 October 2015

The likelihood function $L(\mu) = \left[1 + N\left((\mu - \overline{x}))/s\right)^2\right]^{-n/2}$ is derived, where μ is the true value, $\overline{x}$ is the mean, and s^2 is the variance of N measurements. This form approaches a normal for n large, but can be used also for n small. The use of this formula in data modeling is discussed.

1 The Correction

The 2014 paper with this same title (Miller, 2014) unfortunately contains an elementary and embarrassing error in Eq. 1. This equation gives the assumed normal distribution of measurement value x_i as

$$P(x_i|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right) \quad , \tag{1}$$

for given mean μ and standard deviation σ. The original equation is not even dimensionally correct. Using the same analysis as in the original paper, but with this new Eq. 1 as the starting point, the result quoted in the abstract is obtained, which differs from that of the original paper in having the power $n/2$ rather than $n/4$. This new result approaches a normal as $n \to \infty$, as it must.

The mean and variance are defined as follows (note n rather than $n-1$ in the formula for the variance)

*Corresponding author: guthriemiller@gmail.com

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

$$\begin{aligned}\overline{x} &\equiv \frac{1}{n}\sum_{i=1}^{n} x_i \\ s^2 &\equiv \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 \quad .\end{aligned} \tag{2}$$

2 Data modeling

In empirical data modeling based on replicate measurements, a normal (or lognormal) distribution is usually assumed based on the observed mean and variance of the replicate measurements. This is known to be correct as long as the number of measurements is sufficiently large. But how large? This question is left unanswered but is very important, both in experimental design and the interpretation of the experiment results.

The problem with the conventional approach for small n is that the large uncertainty in the estimation of the standard deviation using just a few replicate measurements is not taken into account.

The new formula depends on the number of replicates, as one might guess the correct result should, and approaches a normal in the limit. It can be used even for only 2 replicates ($n = 2$).

As a matter of convention the likelihood function may be assumed to be of the form $\exp(-\chi^2/2)$ as for the normal case. Then, for each measurement point the χ^2 function is given by

$$\chi^2(\mu) = n \log\left[1 + \frac{1}{n}\left(\frac{(\mu - \overline{x}))}{s}\right)^2\right] \quad , \tag{3}$$

where μ is the unknown true value, which would be calculated from the forward model. For independent measurements, the sum of χ^2 over all measurements points can be minimized by varying forward-model parameters. The quality of the final fit can be judged by comparing the final value of the sum of χ^2 to the number of measurement points, or perhaps even for a nonlinear fit to the number of degrees of freedom of the fit, which is the number of data points minus the number of fit parameters. If the final summed χ^2 is significantly greater than the number of measurement points, this indicates a problem, while $\chi^2/N \approx 1$ or less, where N is the number of measurement points, indicates a satisfactory fit. "Error" or "uncertainty" bars on the data points can be defined as points where $\chi = 1$.

3 Lognormal versus normal

Before calculating the mean and variance of the replicate measurements using Eq. 2, a non linear transformation might be applied to the measurement results x_i, for example,

$x_i \to x_i^2$ or $x_i \to \log(x_i)$. The choice of a possible transformation ultimately depends on observation of the actual distribution of measurement results in controlled experiments. In many situations, a lognormal is observed–probably related to the central-limit theorem and the fact that the measurement result is the product of many factors. However, even if a lognormal is generally observed, for a particular measurement point some of the replicates may have negative values, ruling out a log transformation. In that case untransformed variables would need to be used. In measurements involving counting (for example, measurements of radioactivity), negative results sometimes occur in situations where the true result is of the same magnitude as the counting uncertainty, because of the sample counts being less than background.(Miller, 2015)

References

Miller, Guthrie (2014). A Simple and Conservative Empirical Likelihood Function. *Electronic Journal of Applied Statistical Analysis*, 7(2), pages 344–349. (2014).

Miller, Guthrie (2015). *Probabilistic Interpretation of Data–A Physicist's Approach.* Lulu Publications.