Banat's University of Agricultural Sciences and Veterinary Medicine from Timisoara,

Contact: web: http://dse.usab-tm.ro, e-mail: bjb@usab-tm.ro

CONFIDENCE BAND APPROXIMATION USING INTERVAL ARITHMETIC

Simon Brown[1], Kevin C. Pedley[2], Noorzaid Muhamad[3], David C. Simcock[2]

[1] School of Human Life Sciences, University of Tasmania,Launceston, Tasmania 7250, Australia, Simon.Brown@utas.edu.au
[2] Institute of Food, Nutrition and Human Health, Massey University, Palmerston North, New Zealand
[3] Universiti Kuala Lumpur, Royal College of Medicine Perak, 30450 Ipoh, Perak, Malaysia

Abstract. Confidence bands are commonly obtained from linear regression, but they are rarely shown for nonlinear functions. In part this is due to mathematical complexity. A simple, intuitive and easily automated alternative is to employ interval arithmetic to approximate the confidence band. We illustrate the method by applying it to a straight line, and then apply it to the rate equation of a Michaelis-Menten enzyme and a general polynomial. In each case the approximation is generally larger (and never smaller) than the corresponding standard confidence band, but in at least some instances the upper bound of the discrepancy is about 40%.

Keywords**:** confidence interval, interval arithmetic, Michaelis-Menten kinetics, polynomial.

Introduction

Most experiments yield parameter estimates in the form of <*x*> ± ε, where <*x*> is an average and ε is the standard deviation, standard error or other measure of the error of the estimate. While this makes the precision of the estimates clear, the significance of the error can be more difficult to assess. For example, a function $y = f(x; a_i)$, $i = 1, 2, \ldots n$, might have an error associated with each of the n parameters of ε_i, so an estimate of the total error of y (ε_y) would be

$$\varepsilon_y = \sqrt{\sum_{i=1}^{n} \left(\frac{\partial f(x; a_i)}{\partial a_i} \right)^2 \varepsilon_i^2} \ . \qquad (1)$$

Considering each a_i in isolation can be misleading, especially for nonlinear functions. However, numerical differentiation is ill-posed in the sense that small variations in a_i can result in large differences in the computed derivative [LU & PEREVERZEV 2006, CHENG et al. 2007] which may make the application of (1) problematic [BROWN et al. 2012]. An alternative simple and intuitively attractive approach to the problem would be to estimate (1) using interval arithmetic [Moore 1979], which relies on simple algebraic analysis, some of which might be automated [HICKEY et al. 2001, JAULIN et al. 2001], and does not requires differentiation. Here we show that the error estimate obtained using interval arithmetic is only about 40% larger than that estimated from (1) for sufficiently large x.

We summarise the basic operations of interval arithmetic which we then apply to the approximation of the confidence band of a straight line, to illustrate the approach used. We then extend this approach to the rectangular hyperbola used in enzyme kinetics [Briggs & Haldane 1925] and in so many other contexts and to a general polynomial.

Interval arithmetic

Assuming that X = <*x*> ± ε, then the effective lower and upper limits of X are x_L = <*x*> − ε and x_U = <*x*> + ε, respectively, and of course $x_L \le x_U$. This interval can be written as $X = [x_L, x_U]$ and a real number is a degenerate interval in which $x_L = x_U$. If $A = [a_L, a_U]$ and $B = [b_L, b_U]$ are intervals defined in this way, then the following basic arithmetic operations can be defined [Moore 1979]

$$A + B = [a_L + b_L, a_U + b_U]$$

$$A - B = [a_L - b_U, a_U - b_L]$$

$$AB = \begin{bmatrix} \min(a_L b_L, a_L b_U, a_U b_L, a_U b_U), \\ \max(a_L b_L, a_L b_U, a_U b_L, a_U b_U) \end{bmatrix}$$

which reduces to $AB = [a_L b_L, a_U b_U]$ if $0 < a_L < a_U$ and $0 < b_L < b_U$, and

$$\frac{A}{B} = [a_L, a_U]\left[\frac{1}{b_U}, \frac{1}{b_L}\right],$$

if $0 \notin \{b_L, b_U\}$.

In addition, we make use of the width of the interval A, which is given by

$$w(A) = |a_U - a_L| \qquad (2)$$

and

$$A^* = \frac{a_L + a_U}{2}, \qquad (3)$$

Available on-line at http://dse.usab-tm.ro
Banat's Journal of Biotechnology
2012, III(5),

which is the midpoint of the interval A.

Theory

For a straight line $y = ax + c$ for which the coefficient estimates are $\hat{a} = a \pm \varepsilon_a$ and $\hat{c} = c \pm \varepsilon_c$, then $\hat{y} = y \pm \varepsilon_y$ where (1) yields

$$\varepsilon_y = \sqrt{\left(\frac{\partial y}{\partial a}\right)^2 \varepsilon_a^2 + \left(\frac{\partial y}{\partial c}\right)^2 \varepsilon_c^2} = \sqrt{\varepsilon_a^2 x^2 + \varepsilon_c^2} \qquad (4)$$

where x is usually taken in relation to the midpoint of the range (x_{mid}) [Dudewicz & Mishra 1988]. Interval analysis of the equation, based on the same coefficient estimates, implies that $A = [a_L, a_U] = [\hat{a} - \varepsilon_a, \hat{a} + \varepsilon_a]$ and $C = [c_L, c_U] = [\hat{c} - \varepsilon_c, \hat{c} + \varepsilon_c]$, and yields

$$Y = Ax + C = [a_L x + c_L, a_U x + c_U] \qquad (5)$$

for $x > 0$, and so the limits are a linear function of x and the width (2) and midpoint (3) of the interval are

$$w(Y) = |(a_U - a_L)x + c_U - c_L| = 2(\varepsilon_a x + \varepsilon_c) \qquad (6)$$

and

$$Y^* = \frac{1}{2}((a_L + a_U)x + c_L + c_U), = \hat{a}x + \hat{c} = \hat{y} \qquad (7)$$

respectively. In this case the error is symmetrical about $\hat{y} = Y^*$ since the error estimated from interval arithmetic (ξ_y) is

$$\xi_y = \hat{y} - y_L = y_U - \hat{y} = \varepsilon_a x + \varepsilon_c . \qquad (8)$$

Using (4) and (8) and assuming that $x \geq 0$ yields an expression for the error estimated from interval arithmetic (ξ_y) in terms of the total error (ε_y)

$$\xi_y^2 = \left(1 + \frac{2\varepsilon_a \varepsilon_c x}{\varepsilon_a^2 x^2 + \varepsilon_c^2}\right)\varepsilon_y^2 . \qquad (9)$$

From (9) it is clear that (i) $\varepsilon_y^2 \leq \xi_y^2 \leq 2\varepsilon_y^2$ for $x \geq 0$ and (ii) $\varepsilon_y^2 = \xi_y^2$ when $x = 0$ or $x \to \infty$. Equation (9) can be used to estimate the discrepancy between ξ_y and ε_y

$$\Delta = \frac{\sqrt{\xi_y^2 - \varepsilon_y^2}}{\varepsilon_y}, \qquad (10)$$

which has a maximum at ($\varepsilon_c/\varepsilon_a$, 1). Equation (10) provides an upper bound on the relative difference between ξ_y and ε_y $\left((\xi_y - \varepsilon_y)/\varepsilon_y\right)$, since

$$\xi_y^2 - \varepsilon_y^2 = (\xi_y + \varepsilon_y)(\xi_y - \varepsilon_y) > (\xi_y - \varepsilon_y)^2 .$$

Combining these bounds yields

$$1 \geq \Delta \geq (\xi_y - \varepsilon_y)/\varepsilon_y \geq 0 \qquad (11)$$

which can be improved using the empirical approximation

$$(\xi_y - \varepsilon_y)/\varepsilon_y \approx (\sqrt{2} - 1)\Delta \qquad (12)$$

(Figure 1) which facilitates estimation of ε_y from ξ_y.

The rate of a reaction catalysed by a Michaelis-Menten enzyme depends on the concentration of substrate (S)

$$v = \frac{V_{\max} S}{K_m + S}, \qquad (13)$$

where $V_{\max} \pm \varepsilon_{Vm}$ is the maximum rate and $K_m \pm \varepsilon_{Km}$ is a measure of the affinity of the enzyme for the substrate. Using (1), the confidence band of (13) is

$$\varepsilon_v = v\sqrt{\left(\frac{\varepsilon_{Vm}}{V_{\max}}\right)^2 + \left(\frac{\varepsilon_{Km}}{K_m + S}\right)^2} \qquad (14)$$

[Brown et al 2012]. Interval analysis yields

$$V = \left[\frac{V_{\max L} S}{K_{mU} + S}, \frac{V_{\max U} S}{K_{mL} + S}\right] \qquad (15)$$

which means that the bounds are not symmetrical because

$$\frac{v_U - v}{v} = \frac{\varepsilon_{Vm}}{V_{\max}} + \frac{\varepsilon_{Km}(V_{\max} + \varepsilon_{Vm})}{V_{\max}(K_m - \varepsilon_{Km} + S)}$$

and

$$\frac{v - v_L}{v} = \frac{\varepsilon_{Vm}}{V_{\max}} + \frac{\varepsilon_{Km}(V_{\max} - \varepsilon_{Vm})}{V_{\max}(K_m + \varepsilon_{Km} + S)}$$

are not equal. If $\varepsilon_{Km} << K_m + S$ and $V_{\max} \pm \varepsilon_{Vm} \approx V_{\max}$ then

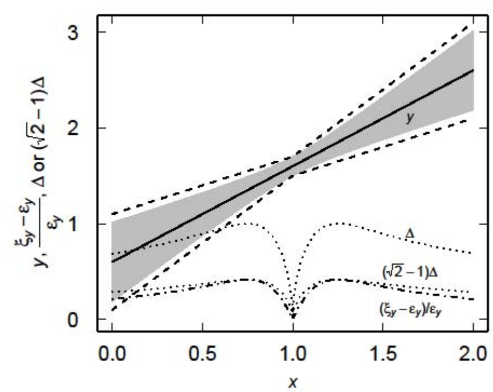

Figure 1. Confidence band estimates for a straight line ($a = 1$, $\varepsilon_a = 0.4$, $c = 0.3$, $\varepsilon_c = 0.1$). The grey region represents the confidence

band estimated from (4) and the dashed curves represents that estimated using (8). The dotted curves are given by (10) and (12) and the lowest curve is calculated from (4) and (8).

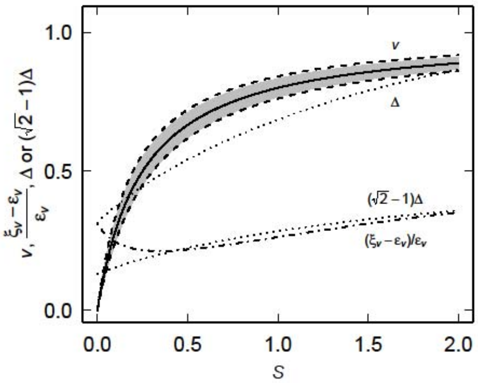

Figure 2. Confidence band estimates for the rate of reaction of a Michaelis-Menten enzyme (13) assuming $V_{max} = 1$, $\varepsilon_{Vm} = 0.01$, $K_m = 0.25$ and $\varepsilon_{Km} = 0.05$. The grey region represents the confidence band estimated from (14) and the dashed curves represents that estimated using (15). The dotted curves are given by (10) and (12) and the lowest curve is calculated from (14) and (16).

$$\xi_v = v\left(\frac{\varepsilon_{Vm}}{V_{max}} + \frac{\varepsilon_{Km}}{K_m + S}\right) \approx v - v_L \approx v_U - v \quad (16)$$

from which

$$\xi_v^2 = \left[1 + \frac{2\left(\frac{\varepsilon_{Vm}}{V_{max}} \frac{\varepsilon_{Km}}{K_m + S}\right)}{\left(\frac{\varepsilon_{Vm}}{V_{max}}\right)^2 + \left(\frac{\varepsilon_{Km}}{K_m + S}\right)^2}\right]\varepsilon_v^2 \qquad (17)$$

and $\xi_v > \varepsilon_v$ since $S \geq 0$. While (17) is of a form similar to (9), there are no values of $S \geq 0$ for which $\xi_v = \varepsilon_v$. However, Δ has a maximum at ($\varepsilon_{Km}V_{max}/\varepsilon_{Vm} - K_m$, 1) and, for sufficiently large S, (11) also applies in this case (Figure 2). So, as for the straight line, (10) provides an upper bound on the relative difference between ξ_y and ε_y $\left((\xi_y - \varepsilon_y)/\varepsilon_y\right)$ and this can be improved using (12) for sufficiently large S *(Figure 2)* which facilitates estimation of ε_y from ξ_y.

For a general polynomial in $x \geq 0$

$$f(x) = \sum_{i=0}^{n} a_i x^i ,$$

each of the coefficients (a_i) of which has an error estimate (ε_i), the total error of $f(x)$ estimated from (1) is

$$\varepsilon_f = \sqrt{\sum_{i=0}^{n} x^{2i}\varepsilon_i^2} \qquad (18)$$

and the interval expression for the polynomial is

$$F = \left[\sum_{i=0}^{n}(a_i - \varepsilon_i)x^i, \sum_{i=0}^{n}(a_i + \varepsilon_i)x^i\right]$$
$$= \left[1 - \frac{\sum_{i=0}^{n}\varepsilon_i x^i}{f(x)}, 1 + \frac{\sum_{i=0}^{n}\varepsilon_i x^i}{f(x)}\right] f(x)$$

from which

$$\xi_f = f(x) - F_L = F_U - f(x) = \sum_{i=0}^{n}\varepsilon_i x^i . \qquad (19)$$

Using (18), (19) can be written in the same form as (9) and (17)

$$\xi_f^2 = \left(1 + \frac{2\sum_{i=0}^{n-1}\sum_{j=i+1}^{n}\varepsilon_i\varepsilon_j x^{i+j}}{\sum_{i=0}^{n}\varepsilon_i^2 x^{2i}}\right)\varepsilon_f^2 \qquad (20)$$

from which it is clear that $\varepsilon_y^2 = \xi_y^2$ when $x = 0$ or $x \longrightarrow \infty$.

Discussion

We have outlined a simple, intuitive and easily automated means of estimating the confidence band of nonlinear functions. While $\xi_y > \varepsilon_y$ in each of the three examples considered here, as is apparent from (9), (17) and (20), the discrepancy (Δ) is not constant. However, for a straight line or the Michaelis-Menten equation Δ is no more than 1 and provides an approximation of the relative difference between ξ_y and ε_y when scaled appropriately (12) as is apparent from *Figures 1 and 2*. We have not defined an apparent upper bound in the case of a general polynomial, but we conjecture that Δ is of a similar magnitude (11).

In those cases where several coefficients or parameters are estimated, it can be misleading to consider one of them while neglecting the others [BROWN ET AL 2012]. The approach outlined here provides an estimate of the confidence band using only simple

Available on-line at http://dse.usab-tm.ro
Banat's Journal of Biotechnology
2012, III(5),

algebra. Obviously, where the statistical significance of a difference in small it is important to carry out the full analysis, but in many cases the slight over-estimate of the confidence band given by interval arithmetic may be sufficient.

References

1. Lu, S.; Pereverzev, S.V., Numerical differentiation from a viewpoint of regularization theory, *Mathematics of Computation* **2006**, *75*, 1853-1870.
2. Cheng, J.; Jia, X.Z.; Wang, Y.B., Numerical differentiation and its applications, *Inverse Problems in Science and Engineering* **2007**, *15*, 339-357.
3. Brown, S.; Muhamad, N.; Pedley, K.C.; Simcock, D.C., A simple confidence band for the Michaelis-Menten equation, *International Journal of Emerging Sciences* **2012**, in press,
4. Moore, R.E., *Methods and applications of interval analysis*, Society for Industrial and Applied Mathematics, Philadelphia, 1979.
5. Hickey, T.; Ju, Q.; van Emden, M.H., Interval arithmetic: from principles to implementation, *Journal of the Association of Computing Machinery* **2001**, *48*, 1038-1068.
6. Jaulin, L.; Kieffer, M.; Didrit, O.; Walter, E., *Applied interval analysis*, Springer-Verlag, London, 2001.
7. Briggs, G.E.; Haldane, J.B.S., A note on the kinetics of enzyme action, *Biochemical Journal* **1925**, *19*, 338-339.
8. Dudewicz, E.J.; Mishra, S.N., *Modern mathematical statistics*, John Wiley and Sons, Inc., New York, 1988.

Received: April 28, 2012
Accepted: May 12, 2012