

Avoiding model selection bias in small-sample genomic datasets

Daniel Berrar, Ian Bradbury and Werner Dubitzky

Bioinformatics (2006) **22**(10), 1245–1250

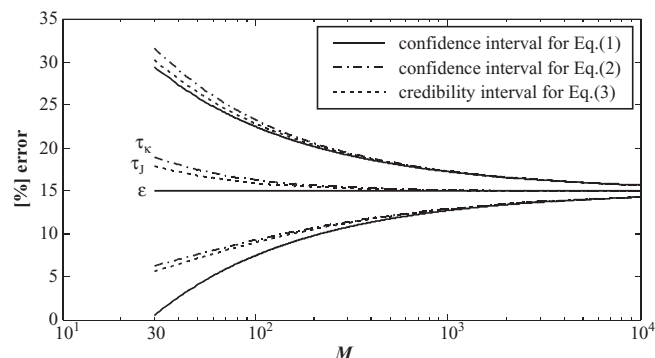
The authors would like to apologize for an error in the calculation of the true error rate τ based on Equation (1), which entailed too narrow intervals for this equation.

On page 1248, the confidence interval for τ based on Equation (1) is given by $\tau = 15.00 \pm 0.08$. Here, the center is given as percentage (i.e. 15%), whereas the limits are not. The correct approximate interval is $\tau \approx 0.15 \pm 0.07499$. As a consequence, the statements concerning the comparisons of the intervals based on the different equations need to be corrected:

For $M = 100$ and $\varepsilon = 0.15$, the widths of the intervals are 0.14998 for Equation (1), 0.13978 for Equation (2) and 0.13826 for Equation (3).

For $M = 300$ and $\varepsilon = 0.15$, the corresponding widths are 0.08415 for Equation (1), 0.08079 for Equation (2) and 0.08048 for Equation (3). The confidence interval based on Equation (1) as

shown in Figure 1 is too narrow. The corrected figure is shown below.



Page 1247, second column, lines 19–20: ‘[...] an approximate $(1 - \alpha)\%$ -credibility interval [...]’ should be ‘[...] an approximate $(1 - \alpha)$ -credibility interval [...]’.