# Abstract


**Testing Hypotheses About Psychometric Functions**

An investigation of some confidence interval methods, their validity, and their use in the assessment of optimal sampling strategies.

N. Jeremy Hill, St. Hugh's College, University of Oxford, UK.
D. Phil. Thesis, Trinity Term 2001.

Various methods for computing confidence intervals and confidence regions for the threshold and slope of the psychometric function were investigated in the context of block-design psychophysical experiments of the sort that are typically carried out with trained adult human observers.

Several variations on the bootstrap method, along with the more traditional methods of probit analysis, were tested using computer simulation, comparing (a) the accuracy of overall coverage, (b) the *balance* of coverage between the two sides of a two-tailed interval, and (c) the *stability* of coverage with regard to variation in the total number of observations and in the distribution of stimulus values. For thresholds, the bootstrap percentile and bias-corrected accelerated ($BC_a$) methods were the most reliable, and for slopes the $BC_a$ method was generally the best choice. The differences between methods were greater, and their performance was generally poorer, (a) for slopes than for thresholds, (b) in the two-alternative forced-choice than in the yes-no design, and (c) when the observer's rate of guessing and/or "lapsing" cannot be assumed to be zero and must therefore be estimated. The problem of bias in the initial slope estimate was also exacerbated by the addition of guessing and lapsing rates as nuisance parameters.

Computer-intensive confidence interval methods were also used to assess the relative efficiency of different distributions of stimulus values, with regard to the estimation of threshold and slope. The most efficient sampling patterns shared certain characteristics irrespective of the number of blocks into which they were divided. Certain unevenly spaced sampling patterns were marginally more efficient than evenly spaced ones.

Further simulations illustrated that, given broad assumptions about the way in which stimulus intensities are chosen in realistic experiments, the assumption of fixed stimulus values, which is intrinsic to the bootstrap methods commonly applied to psychometric functions, may lead to low coverage.