New Developments and Applications in Experimental Design
IMS Lecture Notes – Monograph Series (1998) Volume 34

OPTIMAL DESIGN FOR ITEM CALIBRATION IN COMPUTERIZED ADAPTIVE TESTING: THE 2PL CASE

BY STEVEN G. BUYSKE[1]

Rutgers University

Item response theory is the psychometric model used for standardized tests such as the Graduate Record Examination. A test-taker's answer on an item is modelled as a binary response with success probability depending on parameters for the test-taker and the item. The advent of computerized adaptive versions of these tests leads to sequential design problems. We show how the need for estimation of the item parameters with their ultimate use in mind leads to a locally L-optimal design criterion. A sequential implementation of the optimal design is presented, which is 52% more efficient than the most common current design.

1. Introduction. The basis for modern standardized testing is known as Item response theory [Hambleton and Swaminathan (1985)]. The key idea is that each test question, generally called an item, is characterized by a few parameters, and each test-taker is characterized by a single parameter, generally called proficiency or ability. The probability that a given test-taker answers a given item correctly is given by a function of both the item's and the test-taker's parameters. Conditional on those parameters, the response on one item is independent of the responses to other items.

The model that we focus on in this paper is

$$P(\text{correct response} \mid a, b, \theta) = \frac{1}{1+\exp(-a(\theta-b))}.$$

Here θ is the test-taker's proficiency, and a and b are item parameters. Generally a is called the discrimination and b the difficulty parameter. This model is known as the two parameter logistic, or 2PL, model. Another popular model, the 3PL model, includes a non-zero left asymptote, and is treated in Buyske (1998).

Historically, essentially equivalent paper-and-pencil tests were given at a fixed time to an extremely large number of people. In that case, the item and test-taker parameters could be jointly estimated by maximum likelihood methods. Current interest in

---

[1] Research supported in part by an Educational Testing Service fellowship.
Received September 1997; revised April 1998.
AMS 1991 subject classifications. Primary 62K05; Secondary: 62L05, 62L12, 62P15.
Key words and phrases. Item response theory, online calibration, sequential design.