The Conditional Distribution of the sum of Independent and non Identically Distributed Binary Answers in Item Response Models

Jorge González*
Pontificia Universidad Católica de Chile, Santiago, Chile - jgonzale@mat.puc.cl

Marie Wiberg
Umeå University, Umeå, Sweden - marie.wiberg@stat.umu.se

Alina A. von Davier
Educational Testing Service, Princeton, NJ, United States - avondavier@ets.org

In the field of educational measurement, item response theory (IRT) models specify the probability of correctly answering an item on a test, conditionally on ability. In the case of binary scored items (when an answer is either right or wrong), the information provided by an individual's response pattern is summarized with a test score which typically corresponds to the summation of correct answers. Knowing the distribution of test scores is crucial for the implementation of various psychometric methods which are necessary to ensure accurate and fair score reports leading to good decisions.

Using item response theory (IRT), the frequency distribution of test scores conditional on a given ability can be obtained. If the probability of answering correct is constant for all items on a test, the problem is easy as the resulting distribution is the Binomial. For the more realistic case when scores are the summation of the correct responses, and when the probability of a correct answer varies from item to item, an explicit form of the distribution is more difficult to obtain.

The aim of this paper is to propose the use of the Poisson-Binomial distribution to model the conditional distribution of test scores in an IRT framework. The Poisson-Binomial is the probability distribution of the sum of binary random variables in the independent but non identically distributed, and it has been used in a variety of fields such as reliability modeling and decision theory. To the best of our knowledge, the Poisson Binomial distribution has not yet been used in psychometrics.

A discussion about the the use of this distribution in IRT True-score equating, IRT observed-score equating and local equating is included. Numerical examples with real data as well as simulations are used in order to illustrate the potentials of using the distribution in the test measurement context.

Keywords: conditional scores distribution; item response theory; test equating; non identically distributed case.