A Note on the Poisson's Binomial Distribution in Item Response Theory

Applied Psychological Measurement
2016, Vol. 40(4) 302–310
© The Author(s) 2016
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0146621616629380
apm.sagepub.com

Jorge González¹, Marie Wiberg², and Alina A. von Davier³

Abstract

The Poisson's binomial (PB) is the probability distribution of the number of successes in independent but not necessarily identically distributed binary trials. The independent non-identically distributed case emerges naturally in the field of item response theory, where answers to a set of binary items are conditionally independent given the level of ability, but with different probabilities of success. In many applications, the number of successes represents the score obtained by individuals, and the compound binomial (CB) distribution has been used to obtain score probabilities. It is shown here that the PB and the CB distributions lead to equivalent probabilities. Furthermore, one of the proposed algorithms to calculate the PB probabilities coincides exactly with the well-known Lord and Wingersky (LW) algorithm for CBs. Surprisingly, we could not find any reference in the psychometric literature pointing to this equivalence. In a simulation study, different methods to calculate the PB distribution are compared with the LW algorithm. Providing an exact alternative to the traditional LW approximation for obtaining score distributions is a contribution to the field.

Keywords

Poisson's binomial distribution, compound binomial distribution, Lord and Wingersky's recursive formula, score distributions

Introduction

Lord (1980) pointed out that using item response theory (IRT), the frequency distribution of test scores, X , conditional on a given ability, θ , could be obtained. For the case when X is the summation of the correct responses, Lord recognized that an explicit formula of the conditional distribution was difficult to obtain in a simple form, except for the unrealistic scenario when all items have the same probability of being answered correctly, in which case the binomial distribution appears naturally (see Lord, 1980, Section 4.1). For the non-identically distributed case,

¹Pontificia Universidad Católica de Chile, Santiago, Chile

²Umeå University, Sweden

³Educational Testing Service, Princeton, NJ, USA

Corresponding Author:

Jorge González, Facultad de Matemáticas, Pontificia Universidad Católica de Chile, Casilla 306, Correo 22, Santiago 7820436, Chile.

Email: jorge.gonzalez@mat.uc.cl

that is, when the probabilities of a correct answer differ across items, the use of probability generating functions leads to what is called the compound binomial (CB) distribution, also known as the generalized binomial (Lord, 1980, Section 4.1; Kendall & Stuart, 1958; Lord & Novick, 1968, Section 16.12). The CB distribution can be directly used to obtain the score probabilities, but this practice is computationally demanding for some cases (as it will be demonstrated later).

Wang (1993) presented an explicit form for the distribution of the number of successes in independent but not necessarily identically distributed binary trials, the Poisson's binomial (PB) distribution, and studied many of its properties. Wang pointed out that this distribution has played an important role in probability theory, and it dates back at least to Poisson (1837). Direct calculation of probabilities using the PB distribution suffers from similar computational problems as the CB. Nevertheless, efficient algorithms are available for the estimation of the PB distribution. To our knowledge, neither the PB distribution nor the algorithms to obtain the score probabilities seemed to have been referenced in the psychometric literature.

First, we show that the CB and PB distributions lead to the same probabilities of the number of successes in independent binary trials. Second, we conduct a simulation study to evaluate the performance of these algorithms. Third, the well-known Lord and Wingersky (1984) algorithm is shown to be equivalent to one of the methods used for the estimation of the PB distribution.

Theoretical Models

PB

Let X denote the total number of successes in n independent Bernoulli trials where p_i is the probability of success at trial i . If $p_i = p \forall i$, the distribution of X is the Binomial (n, p) . For the case of independent but non-identically distributed Bernoulli variables, X follows the PB distribution (Wang, 1993). The PB distribution has been used in many areas such as, for example, pool screening (Gao, Aban, & Katholi, 2014), survey sampling (Chen & Liu, 1997), and bioinformatics (Niida, Imoto, Shimamura, & Miyano, 2012).

If p_i is the probability of success at the i th trial, the following defines the probability mass function (PMF) of X (Wang, 1993).

Definition: Let $\mathcal{F}_x = \{A : A \subseteq \{1, \dots, n\}, |A| = x\}$, where $|A|$ denotes the number of elements of A . The PMF of X is

$$f(x; \mathbf{p}) = \Pr(X=x) = \sum_{A \in \mathcal{F}_x} \left(\prod_{i \in A} p_i \right) \left(\prod_{j \in A^c} (1 - p_j) \right), \quad (1)$$

and the corresponding cumulative distribution function (CDF)

$$F_X(x; \mathbf{p}) = \Pr(X \leq x) = \sum_{m=0}^x \sum_{A \in \mathcal{F}_m} \left(\prod_{i \in A} p_i \right) \left(\prod_{j \in A^c} (1 - p_j) \right).$$

To use this distribution in practice, consider $n=3$, in which case the $\mathcal{F}_x$ sets are

$$\mathcal{F}_0 = \{\emptyset\}, \mathcal{F}_1 = \{\{1\}, \{2\}, \{3\}\}, \mathcal{F}_2 = \{\{1, 2\}, \{1, 3\}, \{2, 3\}\}, \mathcal{F}_3 = \{\{1, 2, 3\}\}.$$

To calculate the probability that two successes are obtained, we use Equation 1 to obtain

$$\begin{aligned}
 f(2, \mathbf{p}) &= \left(\prod_{i \in \{1, 2\}} p_i \right) \left(\prod_{j \in \{3\}} (1 - p_j) \right) + \left(\prod_{i \in \{1, 3\}} p_i \right) \left(\prod_{j \in \{2\}} (1 - p_j) \right) \\
 &\quad + \left(\prod_{i \in \{2, 3\}} p_i \right) \left(\prod_{j \in \{1\}} (1 - p_j) \right) \\
 &= p_1 p_2 (1 - p_3) + p_1 p_3 (1 - p_2) + p_2 p_3 (1 - p_1) \\
 &= p_1 p_2 q_3 + p_1 p_3 q_2 + p_2 p_3 q_1,
 \end{aligned}$$

where $q_i = 1 - p_i$. Using similar calculations, the probabilities of obtaining 0, 1, and 3 successes are $f(0, \mathbf{p}) = q_1 q_2 q_3$; $f(1, \mathbf{p}) = p_1 q_2 q_3 + p_2 q_1 q_3 + p_3 q_1 q_2$; $f(3, \mathbf{p}) = p_1 p_2 p_3$, respectively. Note, to calculate the probabilities, all the sets satisfying $|A| = x$ in $\mathcal{F}_x$ must be listed. For instance, with $n = 40$, $\mathcal{F}_{10}$ would contain $\binom{40}{10} = 847,660,528$ elements, so direct calculation of probabilities is not practical for large values of n .

Algorithms for the Calculation of Probabilities

Both approximate and exact alternatives to the direct calculation of probabilities to efficiently obtain the distribution function of the PB model have been proposed and are reviewed here. A well-known approximation used in general statistics is the normal approximation (NA), which is based on the central limit theorem and approximates the CDF of the PB distribution by

$$F_X(x) \approx \Phi\left(\frac{x + 0.5 - \mu}{\sigma}\right), \quad (2)$$

where Φ is the CDF of the standard normal distribution,

$$\mu = E(X) = \sum_{i=1}^n p_i \text{ and } \sigma = [\text{Var}(X)]^{1/2} = \left[\sum_{i=1}^n p_i (1 - p_i) \right]^{1/2}. \quad (3)$$

Not surprisingly, this approximation has also been used in psychometrics for the CB (e.g., Lord & Novick, 1968), which will later be shown to be equivalent to the PB.

An improved version of the NA was described by Volkova (1996; see also Neammanee, 2005) and is known as the refined normal approximation (RNA). Compared with NA, it adds a correction to the skewness of the distribution of X . Under this method, the CDF of the PB is approximated by

$$F_X(x) \approx G\left[\frac{x + 0.5 - \mu}{\sigma}\right], \quad x = 0, 1, \dots, n, \quad (4)$$

where μ and σ are defined as in Equation 3, $G(x) = \phi(x) + \gamma(1 - x^2)\phi(x)/6$, $\phi(x)$ is the probability distribution of the standard normal distribution and

$$\gamma = \sigma^{-3} \sum_{i=1}^n p_i (1 - p_i) (1 - 2p_i).$$

Another approximate method referred here to as the Poisson approximation (PA) is based on a famous inequality established by Le Cam (Le Cam, 1960; Steele, 1994) and uses the Poisson distribution, with μ as defined in Equation 3, to approximate the PMF of the PB by

$$f(x) \approx \frac{\mu^x \exp(-\mu)}{x!}. \quad (5)$$

Among the exact methods, one algorithm for computing the distribution function of the PB was proposed in Fernandez and Williams (2010) where polynomial interpolation and the Discrete Fourier Transform (DFT) are used to derive closed-form formulas for the PB's probability distribution function and CDF. Later, Hong (2013) derived the same closed-form expressions in a simpler way. The method is based on the application of the DFT to the characteristic function of the PB distribution, and it is accordingly called the DFT-CF method. Using this method, the CDF of the PB distribution can be obtained through

$$F_X(x) = \frac{1}{n+1} \sum_{l=0}^n \sum_{m=0}^x \exp(-i\omega l m) z_l, \quad x=0, \dots, n, \quad (6)$$

where i is the imaginary unit, $\omega = 2\pi/(n+1)$, and z_l is the characteristic function of the PB random variable evaluated at ωl , that is, $z_l = \prod_{i=1}^n [1 - p_i + p_i \exp(i\omega l)]$ (Hong, 2013).

Other exact methods are recursive and initiated as approximate expressions for the calculation of the PB's CDF. In an early work, Walsh (1955) proposed a method based on power expansions of $(p_i - \bar{p})$. Interestingly, Walsh's method was also used in Lord and Novick (1968, Section 23.10) and later in Yen (1984). Successive work regarding the distribution of the number of successes in independent binary trials includes Hoeffding (1956), Darroch (1964), Samuels (1965), and Nedelman and Wallenius (1986). Thomas and Taub (1982) developed a recursive algorithm based on the probability generating function of the random variable X , which is capable to generate exact values of the distribution without explicitly enumerating all the sets satisfying $|A| = x$ in $\mathcal{F}_x$.

The probability generating function of a random variable X taking values $0, 1, 2, \dots$ is defined as $P(s) = \sum_x s^x \Pr(X=x)$ (Grimmett & Welsh, 2014), and its explicit form for the case when X is the number of successes in independent binary trials is known to be

$$P(s) = \prod_{i=1}^n (q_i + p_i s). \quad (7)$$

It follows that one can extract the values $\Pr(X=x)$ grouping the coefficients of s^x in the expansion of Equation 7. Thomas and Taub (1982) used this result and the fact that Equation 7 can be written as an n th degree polynomial in s in the form

$$P_n(s) = \alpha_{n1}s^0 + \alpha_{n2}s^1 + \alpha_{n3}s^2 + \dots + \alpha_{n,n+1}s^n$$

so that $\Pr_n(X=x) = \alpha_{n,x+1}$ can be calculated using the recursive formula (RF)

$$\alpha_{ij} = \alpha_{i-1,j-1}p_i + \alpha_{i-1,j}q_i, \quad (8)$$

with the conditions $\alpha_{11} = q_1$, $\alpha_{12} = p_1$, $\alpha_{i1} = \alpha_{i-1,1}q_i$ if $i \geq 2$, and $\alpha_{i,n+1} = \alpha_{i-1,n}p_i$ if $i \geq 2$. A program written in BASIC implementing this RF is presented in Barlow and Heidtmann (1984). Note that the use of the probability generating function leads directly to the calculation of probabilities. For instance for $n=3$, one has

$$\begin{aligned}
\sum_{x=0}^3 \Pr(X=x)s^x &= \prod_{i=1}^3 (q_i + p_i s) \\
&= (q_1 + p_1 s)(q_2 + p_2 s)(q_3 + p_3 s) \\
&= (q_1 q_2 q_3) + (q_1 q_2 p_3 + q_1 p_2 q_3 + p_1 q_2 q_3)s \\
&= (p_1 p_2 q_3 + q_1 p_2 p_3 + p_1 q_2 p_3)s^2 + (p_1 p_2 p_3)s^3
\end{aligned}$$

where s is used to group terms so that

$$\begin{aligned}
P(X=0) &= q_1 q_2 q_3, P(X=1) = q_1 q_2 p_3 + q_1 p_2 q_3 + p_1 q_2 q_3, \\
P(X=2) &= p_1 p_2 q_3 + q_1 p_2 p_3 + p_1 q_2 p_3, P(X=3) = p_1 p_2 p_3
\end{aligned}$$

which coincide with the probabilities calculated using Equation 1. This strategy is also used in Lord (1980, Section 4.1). Next, we will show that the RF actually corresponds to the Lord and Wingersky (1984) recursive algorithm.

IRT Models

Let $\mathbf{X} = (X_1, \dots, X_n)$ be an individual's response pattern to n binary items where $X_i \in \{0, 1\}$. Let θ denote the ability of the individual, ω_i an item parameter vector, and p_i the probability that the individual answers item i correctly. The underlying probability model is then $X_i | \theta, \omega_i \sim \text{Bernoulli}(p_i(\theta, \omega_i))$, where the probability of success can be defined with different IRT models. Local independence is assumed, which means that, conditionally on θ , the answers of the individual to the n items in the test are independent. Let $X = \sum_{i=1}^n X_i$ be the sum score of an individual with ability θ . Assuming that individuals with a given ability θ will correctly answer to each of n items of a test with probability p_i ($i = 1, \dots, n$), the conditional distribution of X for a given θ is the CB (Lord & Novick, 1968) and is defined as

$$f(x | \theta) = \sum_{T_x} \left(\prod_{i=1}^n p_i^{x_i} q_i^{1-x_i} \right), \quad (9)$$

where T_x is the set containing all the response patterns with exactly x correct answers, and $x = 0, \dots, n$. As an example of its use to calculate the probability of a certain score, let us consider again $n = 3$ items. Then, the set T_x ($x = 0, 1, 2, 3$) contains the following elements:

$$\begin{aligned}
T_0 &= \{(0, 0, 0)\}, T_1 = \{(1, 0, 0), (0, 1, 0), (0, 0, 1)\} \\
T_2 &= \{(1, 1, 0), (0, 1, 1), (1, 0, 1)\}, T_3 = \{(1, 1, 1)\}
\end{aligned}$$

Using Equation 9, for a given ability, the probabilities of earning each of the x values are

$$\begin{aligned}
f(0 | \theta) &= q_1 q_2 q_3, f(1 | \theta) = p_1 q_2 q_3 + q_1 p_2 q_3 + q_1 q_2 p_3, \\
f(2 | \theta) &= p_1 p_2 q_3 + q_1 p_2 p_3 + p_1 q_2 p_3, f(3 | \theta) = p_1 p_2 p_3,
\end{aligned}$$

respectively. As it was the case for $\mathcal{F}_x$, the number of elements in T_x grows rapidly with the number of items, so direct use of the CB is computationally demanding. For instance, with 40 items, to calculate $f(x=10 | \theta)$, the set T_{10} would contain $\binom{40}{10} = 847,660,528$ elements, which complicates the numerical calculations. In the psychometric literature, these difficulties were noted by Yen (1984; see also, Thissen, Pommerich, Billeaud, & Williams, 1995) who

stated that calculating this distribution is laborious because all 2^n response patterns must be evaluated.

As an alternative to the direct calculation of score probabilities, the CB distribution has traditionally been obtained using a recursion formula given by Lord and Wingersky (1984). The formula reads as follows

$$\begin{aligned} f_r(x \mid \theta) &= f_{r-1}(x \mid \theta) q_r, & x=0 \\ &= q_r f_{r-1}(x \mid \theta) + p_r f_{r-1}(x-1 \mid \theta), & 0 < x < r, \\ &= f_{r-1}(x-1 \mid \theta) p_r, & x=r \end{aligned} \quad (10)$$

where $f_r(x \mid \theta)$ is the distribution of sum scores over the first r items for examinees of ability θ and $p_r = p_i$ is the probability defined from the chosen IRT model.

Note, (a) in both the PB and CB distributions, the number of elements in the sets $\mathcal{F}_x$ and $\mathcal{T}_x$, respectively, grows rapidly with the number of trials, making the computation of probabilities computationally inefficient. Despite this, both PB and CB can be directly used to obtain the score probabilities, and we have shown that they generate identical probabilities so that these distributions can be considered equivalent; (b) the LW algorithm in Equation 10 is actually the RF shown in Equation 8 (if $\alpha_{n,x+1}$ is replaced by $f_r(x \mid \theta)$). It is, however, surprising that none of our cited references on the RF for the PB actually appear in the psychometric literature. As PB and CB are equivalent, we will evaluate the described alternative methods to calculate score probabilities next.

Simulation Example

In this empirical example, the R (R Core Development Team, 2014) package *poibin* (Hong, 2013) was used to obtain the CDFs for the PB distribution using all the described methods (i.e., NA, RNA, PA, DFT-CF) in the previous section in addition to the LW recursive algorithm. The test scores were simulated under the two-parameter logistic IRT model. Individuals were sampled from a $N(0, 1)$ distribution and item parameters from $a \sim U(0.1, 1.5)$ and $b \sim N(0, 1)$. Varying factors were the number of test items $n = 10, 30, 50, 80, 100$, and the sample size $N = 1000, 5000, 10000$. The mean absolute error was calculated over 100 replications for each combination of factors and was defined as

$$\text{MAE} = \frac{1}{100} \sum_{l=1}^{100} \left(\sum_{x=0}^n |F(x) - F^*(x)| \right)_l,$$

where $F(x)$ is calculated using either the LW, NA, RNA or the PA algorithm, and $F^*(x)$ is calculated using the DFT-CF method. This method was chosen as criterion as it produces exact values of probabilities and, in comparison with the LW, is not recursive. To calculate both $F(x)$ and $F^*(x)$, item parameter estimates are obtained for each replicated data set in the simulation.

In many practical applications, the marginal score distribution $f(x)$ is of interest (for instance, in IRT observed-score equating); we integrated out the value of θ by averaging over the range $\theta = -2.0, -1.9, \dots, 1.9, 2.0$, although other methods such as quadrature points or estimates of θ can also be used (Kolen & Brennan, 2014, p. 199). A comparison among the actual conditional distributions for three values of θ , $-2.0, 0.0$, and 2.0 , is given in the supplementary material. The conditional distribution of scores is of interest, for instance, in local equating methods (van der Linden, 2011).

Table 1. Mean Absolute Error for the Estimation of Marginal Distributions for Different Test Lengths.

N	LW	NA	RNA	PA
10	0.000	0.029	0.011	0.462
30	0.000	0.020	0.005	0.644
50	0.000	0.015	0.003	0.759
80	0.000	0.012	0.002	0.832
100	0.000	0.011	0.002	0.869

Note. LW = Lord and Wingersky; NA = normal approximation; RNA = refined normal approximation; PA = Poisson approximation.

Results

Table 1 shows the mean absolute error for each combination of the factors in the simulation for the marginal test score distribution $F(x)$. Because increasing the sample size from $N = 1000$ to $N = 10000$ produced no significant differences or pattern, only results for $N = 1000$ are shown. For all cases, the commonly used recursive LW algorithm performs equally well as the exact DFT-CF method, with results agreeing to the ninth decimal place (although only four decimals are shown here). LW can thus be considered a method that accurately estimates the score distributions.

Studying the approximation methods, PA does not perform well for most cases. The RNA performs better than the NA in all cases and both methods improve with the number of items. Interestingly, when the test length increase ($n \geq 30$), the mean absolute error of the RNA method is less than 0.005. This last finding could suggest that the RNA is a competitive method for the estimation of the score distributions as an alternative to the LW recursive algorithm. The results for conditional distributions of test scores are in line with these and are given as supplemental material.

Discussion

We have shown that in the distribution of the number of successes in independent binary trials, the PB distribution is equivalent to the CB distribution seen in psychometrics. The LW recursive algorithm for CBs was shown to be equivalent to an RF derived for the PB distribution. Alternative methods, both approximate and exact, were introduced for the calculation of PB probabilities and evaluated in the IRT framework.

We used four different approaches for the estimation of the test score distributions. Some approximation methods were competitive in the cases where the score distributions were marginalized over θ . Because the marginal distributions of scores $F(x)$ and $F(y)$ of two test forms to be equated, X and Y , respectively, are used in observed-score equating, the approximation methods can be a convenient alternative to the recursive LW algorithm when performing IRT observed-score equating.

For the case of conditional distributions, approximate methods perform worse when extreme values of abilities are considered in comparison with average values of ability. This could have consequences in equating methods that are based on conditional score distribution functions as, for example, local equating. Because a family of equating functions is defined by different values of θ , the equating for individuals with extreme abilities could be distorted. It is, however, necessary to know to what extent the differences seen for the conditional distributions could have an effect on the actual equating function. This is a topic of future research.

There is an advantage of having a compact and exact mathematical definition for the PB distribution as random variables can directly be generated from this model. This could help in

making fairer simulation studies, for instance, when comparing equating methods, as the advantages and disadvantages of using one or another method to simulate score data are not always clear (Sinharay, Holland, & von Davier, 2011). This topic is currently being investigated by the authors. Other potential benefit of knowing the exact mathematical definition of the PB distribution could arise in multistage testing if the conditional test scores of the individuals are used at each stage of testing (Haberman & von Davier, 2014). In this article, local independence is assumed in the specification of IRT models. The psychometric literature, however, has questioned the independence of items in IRT models (Gibbons & Hedeker, 1992; Wainer & Kiely, 1987). Exploring the score distribution in the non-independent, non-identically distributed case is an interesting topic for future research.

Recent work improving the described algorithms for the PB and showing new applications can be found in Barrett and Gray (2014), and in the psychometric field in Cai (2014). How the presented algorithms would extend to the case of polytomous items data (e.g., Thissen et al., 1995) is also a topic of future research.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Jorge González was partially funded by Fondo Nacional de Desarrollo Científico y Tecnológico (FONDECYT) Grant 1150233. The research in this article by Marie Wiberg was funded by the Swedish Research Council Grant 2014-578.

Supplemental Material

The online appendices are available at <http://apm.sagepub.com/supplemental>

References

- Barlow, R. E., & Heidtmann, K. D. (1984). Computing k-out-of-n system reliability. *IEEE Transactions on Reliability*, 33, 322-323.
- Barrett, B., & Gray, J. (2014). Efficient computation for the Poisson binomial distribution. *Computational Statistics*, 29, 1469-1479.
- Cai, L. (2014). Lord-Wingersky algorithm version 2.0 for hierarchical item factor models with applications in test scoring, scale alignment, and model fit testing. *Psychometrika*, 2, 535-559.
- Chen, S. X., & Liu, J. S. (1997). Statistical applications of the Poisson-binomial and conditional Bernoulli distributions. *Statistica Sinica*, 7, 875-892.
- Darroch, J. N. (1964). On the distribution of the number of successes in independent trials. *The Annals of Mathematical Statistics*, 35, 1317-1321.
- Fernandez, M., & Williams, S. (2010). Closed-form expression for the Poisson-binomial probability density function. *IEEE Transactions on Aerospace and Electronic Systems*, 46, 803-816.
- Gao, H., Aban, I., & Katholi, C. (2014). Pool screening: An example of independent non-identical Bernoulli trial. *Communications in Statistics: Simulation and Computation*. Advance online publication. doi:10.1080/03610918.2014.941486
- Gibbons, R. D., & Hedeker, D. (1992). Full information item bi-factor analysis. *Psychometrika*, 57, 423-436.
- Grimmett, G., & Welsh, D. (2014). *Probability: An introduction* (2nd ed.). Oxford, UK: Oxford University Press.

- Haberman, S. J., & von Davier, A. A. (2014). Considerations on parameter estimation, scoring, and linking in multistage testing. In D. Yan, A. A. von Davier, & C. Lewis (Eds.), *Computerized multistage testing: Theory and applications* (pp. 285-300). London, England: Chapman & Hall.
- Hoeffding, W. (1956). On the distribution of the number of successes in independent trials. *The Annals of Mathematical Statistics*, 27, 713-721.
- Hong, Y. (2013). On computing the distribution function for the Poisson binomial distribution. *Computational Statistics and Data Analysis*, 59, 41-51.
- Kendall, M. G., & Stuart, A. (1958). *The advanced theory of statistics (Vol. 1)*. London, England: C. Griffin.
- Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling and linking: Methods and practices* (3rd ed.). New York, NY: Springer.
- Le Cam, L. (1960). An approximation theorem for the Poisson binomial distribution. *Pacific Journal of Mathematics*, 10, 1181-1197.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Lord, F. M., & Wingersky, M. (1984). Comparison of IRT true-score and equipercentile observed-score equatings. *Applied Psychological Measurement*, 8, 453-461.
- Neammanee, K. (2005). A refinement of normal approximation to Poisson binomial. *International Journal of Mathematics and Mathematical Sciences*, 5, 717-728.
- Nedelman, J., & Wallenius, T. (1986). Bernoulli trials, Poisson trials, surprising variances, and Jensen's inequality. *The American Statistician*, 40, 286-289.
- Niida, A., Imoto, S., Shimamura, T., & Miyano, S. (2012). Statistical model-based testing to evaluate the recurrence of genomic aberrations. *Bioinformatics*, 28, 115-120.
- Poisson, S. D. (1837). *Recherches sur la probabilité des jugements en matière criminelle et en matière civile, précédées des règles générales du calcul des probabilités* [Researches into the Probabilities of Judgements in Criminal and Civil Cases]. Paris, France: Bachelier.
- R Core Development Team. (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Samuels, S. M. (1965). On the number of successes in independent trials. *The Annals of Mathematical Statistics*, 36, 1272-1278.
- Sinharay, S., Holland, P. W., & von Davier, A. A. (2011). Evaluating the missing data assumptions of the chain and poststratification equating methods. In A. A. von Davier (Ed.), *Statistical models for test equating, scaling and linking* (pp. 281-296). New York, NY: Springer-Verlag.
- Steele, J. M. (1994). Le Cam's inequality and Poisson approximations. *The American Mathematical Monthly*, 101, 48-54.
- Thissen, D., Pommerich, M., Billeaud, K., & Williams, V. S. (1995). Item response theory for scores on tests including polytomous items with ordered responses. *Applied Psychological Measurement*, 19, 39-49.
- Thomas, M., & Taub, A. (1982). Calculating binomial probabilities when the trial probabilities are unequal. *Journal of Statistical Computation and Simulation*, 14, 125-131.
- van der Linden, W. J. (2011). Local observed-score equating. In A. A. von Davier (Ed.), *Statistical models for equating, scaling, and linking* (pp. 201-222). New York, NY: Springer.
- Volkova, A. Y. (1996). A refinement of the central limit theorem for sums of independent random indicators. *Theory of Probability and Its Applications*, 40, 791-794.
- Wainer, H., & Kiely, G. (1987). Item clusters and computerized adaptive testing: A case for testlets. *Journal of Educational Measurement*, 24, 185-202.
- Walsh, J. E. (1955). Approximate probability values for observed number of "successes" from statistically independent binomial events with unequal probabilities. *Sankhya: The Indian Journal of Statistics*, 15, 281-290.
- Wang, Y. H. (1993). On the number of successes in independent trials. *Statistica Sinica*, 3, 295-312.
- Yen, W. M. (1984). Obtaining maximum likelihood trait estimates from number-correct scores for the three-parameter logistic model. *Journal of Educational Measurement*, 21, 93-111.