**Logistic Semi-distribution**

**Abstract**

**1. Introduction**

The logistic function is specified by two parameters *a* and *b* which are slope and location, respectively. It is cumulative probability of variable *x* given parameters *a* and *b*, specified by formula 1.1.

|  |  |
| --- | --- |
|  | **(1.1)** |

Where *a* >0 and *exp*(*x*) denotes natural exponent function *ex*.

The function specified by equation 1 characterizes two-parameter logistic (2PL) model where its limit is 0 if *x* approaches negative infinity. Logistic function is applied widely in many scientific domains but this research focuses on mathematical subjects based on logistic function such as item response theory, logistic regression, and neural network. Some mathematicians (Baker & Kim, 2004, pp. 15-18) researched logistic function and they established the relationship between logistic function and normal distribution, specified by formula 1.2.

|  |  |
| --- | --- |
|  | **(1.2)** |

Where Φ(*x*) is cumulative probability function of standard normal distribution.

Author (Nguyen, 2016) calculated mean and variance of logistic function in his research “New version of CAT algorithm by maximum likelihood estimation”. It is asserted that logistic model is a probabilistic distribution which can be called *logistic distribution*. Let Ψ and *ψ* be cumulative probability function (CDF) and probability density function (PDF) of logistic distribution, respectively. In other words, logistic function is the CDF of logistic distribution. Let *μ* and *σ*2 be theoretical mean and variance of logistic distribution, respectively. Formula 1.3 specifies these quantities (Nguyen, 2016, pp. 227-234). Parameters *a* and *b* are called *slope* and *location* of logistic distribution.

|  |  |
| --- | --- |
|  | **(1.3)** |

Formulas 1.1 and 1.3 represent logistic distribution. Note that the logistic distribution had been discovered before I, author (Nguyen, 2016), researched it. I used to think that I discovered it from item response theory. However I calculated the mean and variance of logistic distribution independently without referring its documents such as (Wikipedia, Logistic distribution, 2016) and (Weisstein, Logistic Distribution, n.d.). The CDF Ψ(*x*) and PDF *ψ*(*x*) of logistic distribution are drawn as dot line and solid line in figure 1.1.

**Figure 1.1.** Logistic distribution

As seen in figure 1.1, the CDF Ψ(*x*) reaches highest first-order derivative at point *x*=*b*. This value is *a*/4, which the reason that *a* is called slope parameter.

The point *x*=*b* is also saddle point because the second-order of Ψ(*x*) at *x*=*b* is 0.

The 2nd moment of logistic distribution is:

The 3rd central moment of logistic distribution is (Wikipedia, Logistic distribution, 2016):

It implies:

The 3rd moment of logistic distribution is:

Table 1.1 summarizes important properties of logistic distribution.

**Table 1.1.** Logistic distribution

Let *F*(*x*) be the three-parameter logistic (3PL) function:

Where, .

In item response theory, *c* is called guessing parameter because it is random probability that low-ability examinees guess a correct response to an item when they do not master the item. The function *F*(x) violates an axiom of probability distribution due to:

However, we define that *F*(*x*) is cumulative probability function (CDF) of *logistic semi-distribution*. Let *f*(*x*) be the probability density function (PDF) of logistic semi-distribution.

When *c*=0, logistic semi-distribution becomes logistic distribution as usual. In other words, logistic semi-distribution is the most general one. The CDF *F*(*x*) and PDF *f*(*x*) of logistic distribution are drawn as dot line and solid line in figure 1.2.

**Figure 1.2.** Logistic semi-distribution

It is necessary to determine properties of logistic semi-distribution. The mean of logistic semi-distribution is:

Similarly, the 2nd moment is:

Similarly, the 3rd moment is:

The variance is:

Table 1.2 summarizes important properties of logistic semi-distribution.

**Table 1.2.** Logistic semi-distribution

The next section focuses on parameter estimation of logistic semi-distribution along with its application to item response theory but all estimation methods can be used for 2PL logistic distribution.

**2. Parameter estimation**

Given sample *X* = {*x*1, *x*2,…, *xn*} of size *n*, let and *s*2 be unbiased estimates of *μ* and *σ*2, we apply moment method (Montgomery & Runger, 2003, pp. 229-230) into parameter estimation, as follows:

Let , , and be estimates of *a*2, *b*, and *c*. The estimates are solutions of following equations:

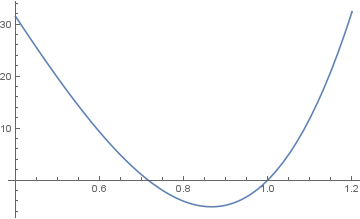
It implies formula 2.1 specifies these estimates as follows:

|  |  |
| --- | --- |
|  | **(2.1)** |

Where,

The last equation in formula 2.1 is called *c*-equation with regard to variable 1–*c*, which is quartic equation. The computational cost to solve it by formal method is expensive. Thus, Newton-Raphson method (Burden & Faires, 2011, pp. 67-69) is the favorite one for solving it. It is necessary to illustrate the estimation formula 2.1. Given three-point sample *X* = {*x*1=1, *x*2=2, *x*3=3}, according to formula 2.1, we have:

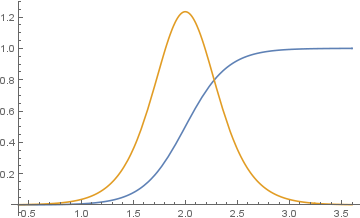
Figure 2.1 depicts the c-equation with regard to 1–*c* given sample *X* = {*x*1=1, *x*2=2, *x*3=3}.



**Figure 2.1.** The *c*-equation with given sample

By using Newton-Raphson method, one solution of *c*-equation is 1–*c* = 1. Therefore, all estimates are determined as follows:

The estimated CDF *F*(*x*) and PDF *f*(*x*) of logistic semi-distribution are drawn as green line and yellow line in figure 2.2.



**Figure 2.2.** Estimated logistic semi-distribution by moment method

Alternatively, the maximum likelihood estimation (MLE) is often used to estimate parameters. Given sample *X* = {*x*1, *x*2,…, *xn*}, the likelihood function of logistic distribution is:

By taking natural logarithm of the *L*(*a*, *b*), we get log-likelihood function of logistic distribution

Where *ln*(*x*) denotes natural logarithm function.

The partial derivatives of *LnL*(*a*, *b*) with regard to *a*, *b*, and *c* are:

The MLE method keeps the estimate fixed because the partial derivative is always less than 0. Thus, without loss of generality, let

The other estimates and are solutions of equations created by setting these partial derivatives to be zero as seen in formula 2.2:

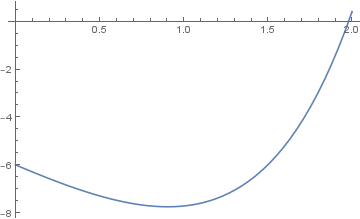
|  |  |
| --- | --- |
|  | **(2.2)** |

Newton-Raphson method (Burden & Faires, 2011, pp. 67-69) is used to solve formula 2.2. It is necessary to illustrate the estimation formula 2.2. Given three-point sample *X* = {*x*1=1, *x*2=2, *x*3=3}, the formula 2.2 becomes:

The value *b*=2 is solution of the second equation due to:

Substituting *b*=2 into the first equation, we have:

Figure 2.3 shows the equation for estimating parameter *a*.



**Figure 2.3.** Equation for estimating parameter *a*

By using Newton-Raphson method manually, the solution of suchequation is *a*=1.98. Therefore, all estimates are determined as follows:

It implies the estimate is:

According to the second equation in formula 2.2, we have:

Let

We have:

The solutions of this trinomial are:

If *c* ≥ 0 then, the solution is removed. If *c* < 0 then, the solution is removed. Figure depicts the equation

Applying Newton-Raphson method (Burden & Faires, 2011, pp. 67-69), we receive the value of *c*.

Later on, we totally determine the estimate as with regard to equation (5) as follows:

**3. Item response theory**

When logistic distribution is applied into Item Response Theory (Wikipedia, Item response theory, 2016), the CDF *ψ* indicates probability of correct response of given examinee whose ability is *x* to an item. In Item Response Theory (IRT), *ψ* is called Item Response Function (IRF) whereas *a* and *b* are called discriminatory parameter and difficult parameter, respectively. Suppose there are *n* items given to an examinee and each item *i* has *ni* optional responses. The number of correct responses of given examinee to item *i* is *ri*. In other words, we have sample *X* = {*r*1, *r*2,…, *rn*}. Author (Nguyen, 2016, pp. 224-227) found out formula 2.3 to calculate the estimate in case of applying logistic distribution to IRT. His finding is derived from the general MLE method with regard to logistic regression models (Czepiel, 2002, pp. 4-5).

|  |  |
| --- | --- |
| Where *x*0 is the examinee’s current ability. | **(2.3)** |

The formula 2.3 is significant finding when it produces a simple calculation for slope estimate whereas Newton-Raphson method must be used to solve complex estimation equation 2.2. However assumption of identical current ability *x*0 over sample *X* in formula 2.3 is only suitable to iterative Computerized Adaptive Testing (CAT) based on IRT in which each iteration focuses only on a new sample and a new current ability. So formula 2.3 is not as comprehensive as formula 2.1 but it is simpler than formula 2.2.

The MLE estimation proposed by formula 2.3 is used for one examinee and many items where each item is subset of binary items. Another MLE method described in (Baker & Kim, 2004, pp. 23-54) is applied in case of many examinees with one item. Suppose there are *n* groups of examinees and each group *i* has *ni* examinees. Among *ni* examinees, there are *ri* ones who give correction responses. All examinees in each group *i* has the same ability *xi*. According to (Baker & Kim, 2004, p. 38), the likelihood function is based on binomial distribution as follows:

By taking natural logarithm of the *L*(*a*, *b*), the log-likelihood function is (Baker & Kim, 2004, p. 38):

The partial derivatives of *LnL*(*a*, *b*) with regard to *a* and *b* are (Baker & Kim, 2004, p. 39):

According to MLE method, estimates and are solutions of equations created by setting these partial derivatives to be zero as seen in formula 2.4 (Baker & Kim, 2004, pp. 27, 39, 47):

|  |  |
| --- | --- |
|  | **(2.4)** |

Newton-Raphson method (Burden & Faires, 2011, pp. 67-69) is used to solve equation 2.4.

Formulas 2.2, 2.3, and 2.4 belong to MLE approach. Formula 2.2 is the most general one. Essentially, formulas 2.2 and 2.4 are similar. Formula 2.3 is the simplest one. Formula 2.4 is the most practical. In other words, formula 2.4 is the most optimal one. However, formulas 2.3 and 2.4 are only applied in IRT. In the most general case, formula 2.1 is used instead because formula 2.2 is too complicated to solve.

According to (Baker & Kim, 2004, p. 55), author Berkson proposed to minimize chi-squared χ2 criterion for estimating IRT parameters *a* and *b* instead of using MLE. This is minimum χ2 estimation, which is also an application of logistic distribution into IRT. Let and be observed frequency and expected frequency of correct response at ability *xi*.

Let and be observed frequency and expected frequency of correct response at ability *xi*.

The χ2 criterion is defined according to formula 2.5, which indicates total deviation between observed frequency and expected frequency over all examinee groups.

|  |  |
| --- | --- |
|  | **(2.5)** |

The smaller the χ2 criterion, the better the estimation is. According to author Berkson, the estimate is minimum point of χ2 criterion. The partial derivatives of χ2 with regard to *a* and *b* are:

Estimates and are solutions of equations created by setting these partial derivatives to be zero (Baker & Kim, 2004, p. 56) as seen in formula 2.6.

|  |  |
| --- | --- |
|  | **(2.6)** |

The Newton-Raphson iterative method (Burden & Faires, 2011, pp. 67-69) can be used to solve equation 2.6 but the author Berkson proposed a transformation technique without iterative process to find out estimates and . Please read (Baker & Kim, 2004, pp. 56-59) for comprehending such excellent technique.

**4. Logistic regression**

**5. Multivariate logistic semi-distribution**

**References**

Baker, F. B., & Kim, S.-H. (2004). *Item Response Theory: Parameter Estimation Techniques* (2nd ed.). CRC Press. Retrieved from https://books.google.com/books?isbn=0824758250

Burden, R. L., & Faires, D. J. (2011). *Numerical Analysis* (9th Edition ed.). (M. Julet, Ed.) Brooks/Cole Cengage Learning.

Czepiel, S. A. (2002). *Maximum Likelihood Estimation of Logistic Regression Models: Theory and Implementation.* Czepiel's website http://czep.net.

Nguyen, L. (2016, April 8). New version of CAT algorithm by maximum likelihood estimation. (B. N. Buszewski, Ed.) *Sylwan Journal, 160*(4), 218-244.

Wikipedia. (2016, March 25). *Item response theory*. (Wikimedia Foundation) Retrieved May 2, 2016, from Wikipedia website: https://en.wikipedia.org/wiki/Item\_response\_theory

Wikipedia. (2016, April 8). *Logistic function*. (Wikimedia Foundation) Retrieved May 2, 2016, from Wikipedia website: https://en.wikipedia.org/wiki/Logistic\_function