Journal of Bibliographical Analyses in Statisctics

CODING

Article

The determination of inflection point on a given ogive curve using the item response theory(IRT) model

Yang Shao[1], Tsair-Wei Chien[2]

[1] Jiaxing Vocational Technical College, Jiaxing, China; yang.shao.ys@outlook.com
[2] Department of Medical Research, Chi-Mei Medical Center, Tainan 700, Taiwan; smile@mail.chimei.org.tw
* Correspondence: smile@mail.chimei.org.tw: +8-866-281-2811

Abstract: The inflection point(IP) on an ogive curve is required for determination using the item response theory(IRT) model. However, we have not seen any article that verified the determination of IP at the location of item difficulty. The way to prove the point at the item difficulty was demonstrated using the differential equation in calculus. We illustrated the data of medical acronyms between 1950 and 2020. The cumulative observed frequencies in Pubmed over the years were used to determine the topic burst by using the IP. Two mathematical models of IRT and the exponential growth curve(=a/(1+be^{-ct}) named EGC model were compared in R^2 and the IPs for each medical acronym. We found that the IRT model outperforms the EGC model. The visual displays were verified in comparison of R^2, and the IPs were demonstrated in this study and provided a reference to future studies regarding the IP determination in other fields, and is not limited to the bibliometric analysis as we did in this study.

Keywords: inflection point, exponential growth model, calculus; ogive curve; IRT; EGC model.

Citation: Shao Y, Chien TW. The determination of inflection point on a given ogive curve using the item response theory(IRT) model . *J Bibliographical Analyses in Statistics* **2021;** *18(3): 31-33*.

Received: 2 July 2021
Accepted: 12 July 2021
Published: 2021/07/21

Introduction

The article published on 16 May 2021, is of interest[1] but remains several questions that are required for clarifications, such as the presentations in Table and Fig. 1 that should be improved further for providing more valuable information to readers because of lacking the topic burst observed in the study. Although improvement has been made in another article[2], the topic burst was not highlighted in the study. Another article[3] displayed topic burst using keywords related to mobile research since 2000. The method for obtaining the year of keywords that was a burst point has not been classified in the study[3] though two references [4,5] were cited. We are thus motivated to cultivate the way to determine the bursting point(namely inflection point(IP)) that appears suddenly at the year when the evolution of keywords or topics is observed.

The item response theory(IRT) model has been applied to determine the IP for countries/regions in COV[6,7]. Nonetheless, the IP determination is still unclear or tedious, not based on a scientific basis. That is, the absolute advantage coefficient(AAC) [8] was applied to determine the IP on a given ogive curve. Whether the calculus that uses the differential equation to solve the problem of the IP location is required to utilization because the IP is located at the IP referred to a point on a smooth plane curve where curvature changes sign from an increasing concave (concave downward) to a decreasing convex (concave upward) shape, or vice versa [9]. In this study,

we aim to verify the IP is located at the point of item difficulty. The medical acronyms will be used to demonstrate the IP determination based on two mathematical models of IRT and the exponential growth curve(=a/(1+be^{-ct}) named EGC model.

Methods

2.1 Data source

We obtained data from the study[1] and extracted eight medical acronyms to illustrate the IP determination and compare the difference in R^2 and the IPs between two mathematic study models.

Two major parts of this study are to verify the results reported in the study[1] and the temporal bar graph used to display the IP and the evolution of keywords(or acronyms).

2.2 The determination of inflection point on a given ogive curve using the second-order derivative in calculus

In the IRT model, the probability can be expressed as Eq.(1):

$$\mathrm{Pro}(X_{ni}|\theta) = \frac{e^{a(\theta_n-\delta_i)}}{1+e^{a(\theta_n-\delta_i)}}, (1)$$

Where parameter a as discrimination parameter is >0 a,b,c are none zero constants,θ, as the t elapsed, is the personal ability,and δ is the item difficulty for n and i, respectively. We simplify Eq.(1) and express it as Eq.(2):

$$f(t) = \frac{e^{a(t-\delta)}}{1+e^{a(t-\delta)}}, (2)$$

The processes of the first and the second-order derivative on t are described below:

$$f'(t) = \left(\frac{e^{a(t-\delta)}}{1+e^{a(t-\delta)}}\right)' = \left[\left(e^{a(t-\delta)}\right)'\left(1+e^{a(t-\delta)}\right) - \left(e^{a(t-\delta)}\right)\left(1+e^{a(t-\delta)}\right)'\right] \div \left(1+e^{a(t-\delta)}\right)^2$$

$$= \left[\left(e^{a(t-\delta)}a\right)\left(1+e^{a(t-\delta)}\right) - a\left(e^{a(t-\delta)}\right)\left(e^{a(t-\delta)}\right)\right] \div (1+e^{a(t-\delta)})^2$$

$$= \frac{(ae^{a(t-\delta)})(1+e^{a(t-\delta)})-a(e^{a(t-\delta)})e^{a(t-\delta)}}{(1+e^{a(t-\delta)})^2} = \frac{(ae^{a(t-\delta)})[(1+e^{a(t-\delta)}-e^{a(t-\delta)})]}{(1+e^{a(t-\delta)})^2}$$

$$= \frac{\left(e^{a(t-\delta)}\right)}{(1+e^{a(t-\delta)})^2}$$

$$f''(t) = \left[f'(t)\right]' = \left\{\frac{\left(e^{a(t-\delta)}\right)}{(1+e^{a(t-\delta)})^2}\right\}'$$

$$= \{-2(ae^{a(t-\delta)}\left[1+e^{a(t-\delta)}\right] - e^{a(t-\delta)}\left[ae^{a(t-\delta)}\right])\} \times \left(1+e^{a(t-\delta)}\right)^{-3}$$

$$= \{-2ae^{a(t-\delta)}\left[1+e^{a(t-\delta)}\right] - e^{a(t-\delta)}ae^{a(t-\delta)}\} \times \left(1+e^{a(t-\delta)}\right)^{-3}$$

Let $e^{a(t-\delta)} = Z$

$$\{-2aZ[1+Z] - aZ^2)\} \times (1+Z)^{-3} = 0$$

Therefore, $(-2aZ - 2aZ^2) \times (1+Z)^{-3} = 0$

$$-2aZ\,[1-Z] \times (1+Z)^{-3} = 0$$

$$\frac{-2aZ\,[1-Z]}{(1+Z)^3} = 0$$

Due to a and Z must be greater than zero, the value of 1-Z must be 0.

Hence, $1-Z=0=1-e^{a(t-\delta)}$

$$e^{a(t-\delta)} = 1$$

$$a(t-\delta) = \ln(1)=0$$

Due to a>0, $t-\delta = 0$ and $t=\delta$

That is, the inflection point on an ogive is at the item difficulty.

2.3 Temporal bar graph used for displaying the evolution of keywords

In tradition, the temporal bar graph(TBG) is just presenting the years on x-axis and bars of counts from high frequency downward the less. No such topic bursts and IP are in the graph. We developed a novel TBG that includes IP, counts in colors with an inner bar, and observed counts with an outer bar, the count in the median with a point to indicate the data pattern(i.e., increasing, decreasing, slow down, and stationarity.

2.3 Estimating model parameters

Model parameters were estimated by using the solver in Microsoft Excel[10,11]. The abstract video was provided in reference[11].

Results

3.1 **Comparison of two mathematic models**

Two points on IPs and R-square are compared in Table 1. We can see that somewhat differences were found in this study, including IPs higher in the EGC model and R-square favoring the IRT model.

Table 1 Comparison of results in IP determinations between the two models

IP	AT IP	start at	end	Rsquare	AAC	Acronym
EGC model						
71	64692	2019	2020	0.00	1.00	COVID
63	21	1977	2020	0.96	0.50	CoV
41	0	1953	2020	0.94	0.81	SARS
71	12210	1950	2020	0.00	0.50	DNA
71	10052	1986	2020	0.15	0.08	HIV
71	9482	1953	2020	0.46	0.50	RNA
71	7088	1954	2020	0.00	0.50	CT
71	6280	1969	2020	0.98	0.49	MRI
IRT model						
71	64692	2019	2020	0.93	1.00	COVID
70	99	1977	2020	1.00	0.5	CoV
71	12290	1953	2020	0.97	0.81	SARS
53	7044	1950	2020	0.99	0.5	DNA
58	7415	1986	2020	0.99	0.5	HIV
55	2604	1953	2020	0.97	0.5	RNA
60	3368	1954	2020	0.96	0.5	CT
62	3389	1969	2020	0.97	0.5	MRI

3.2 **Comparison of visual displays in the two models**

Visualizations are displayed in Figures 1 to 3 in which we can see that the IRT model is superior to the EGV model.

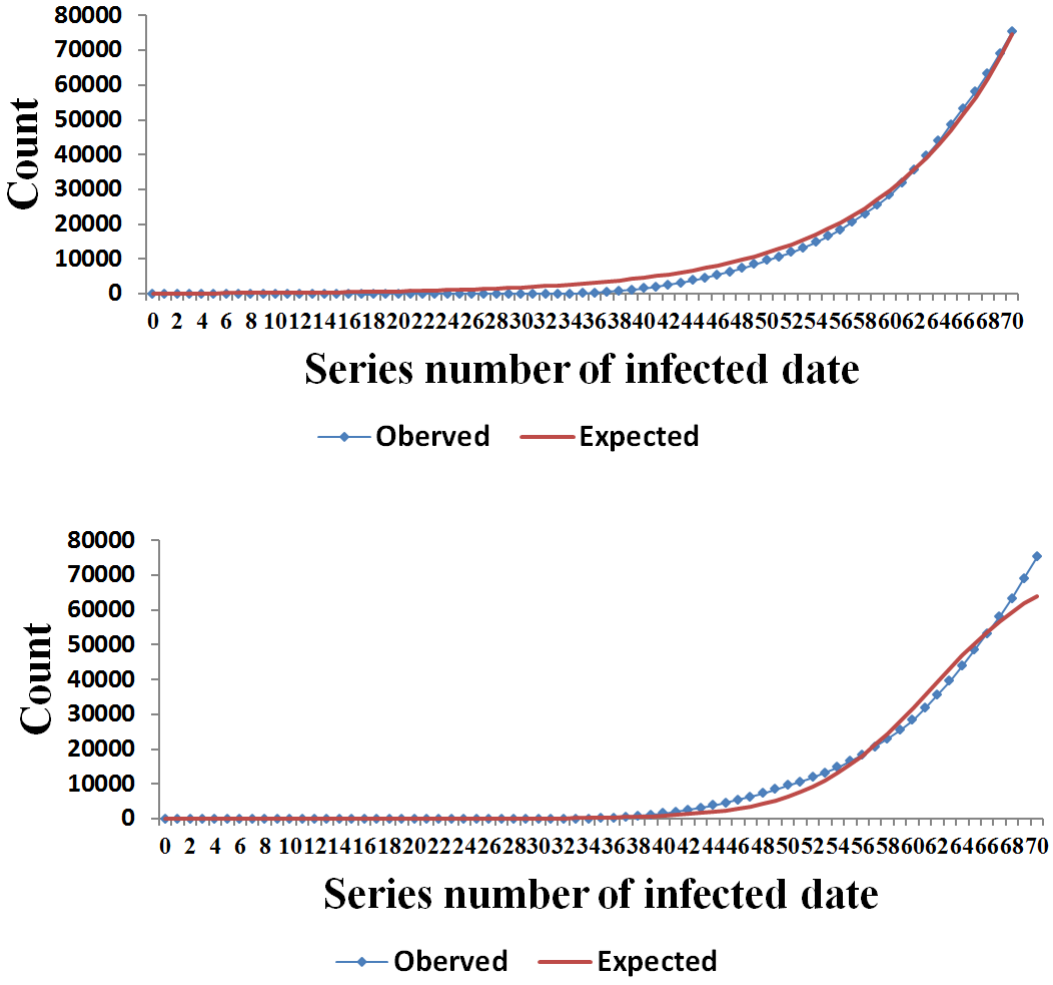


Figure 1 The acronym of MRI in the two models to display(EGV model(top) and IRT model(bottom)).

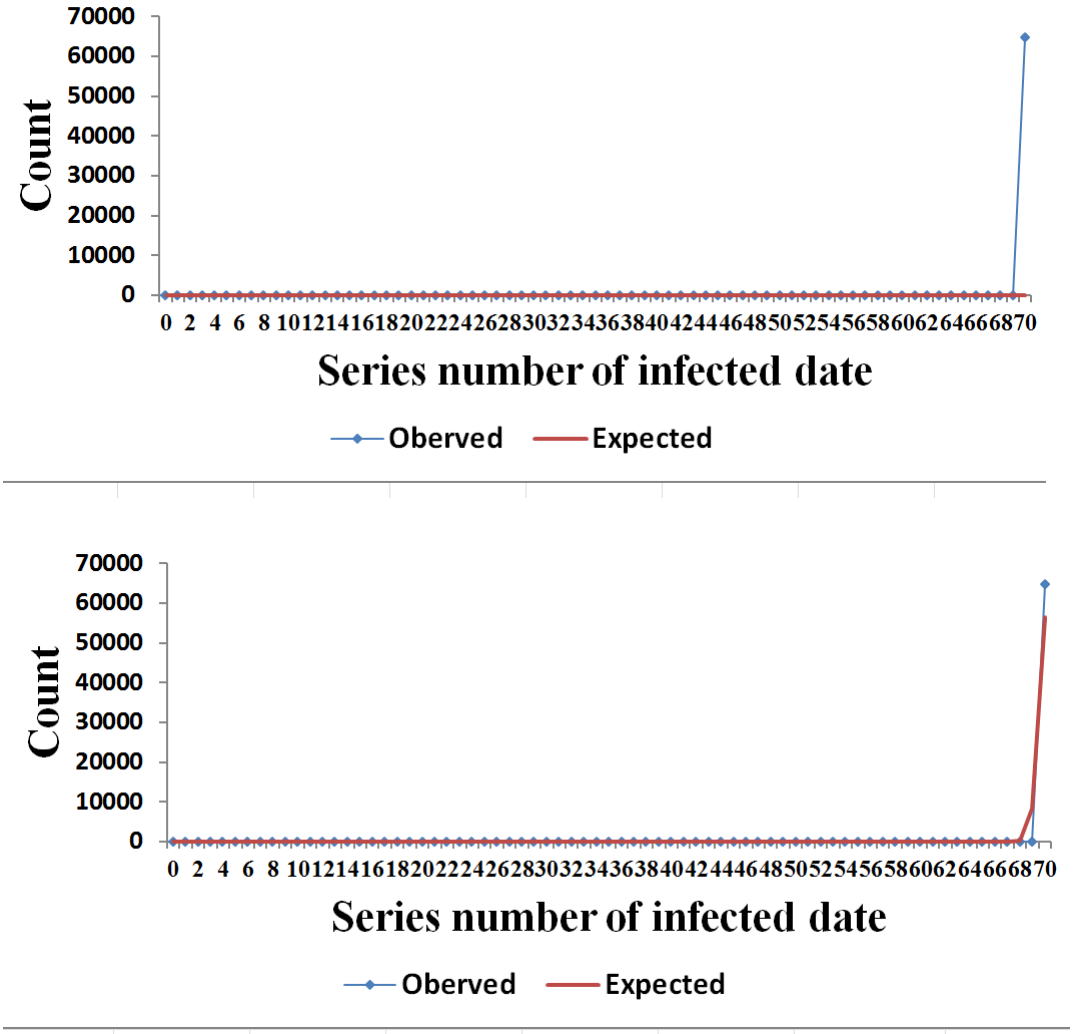
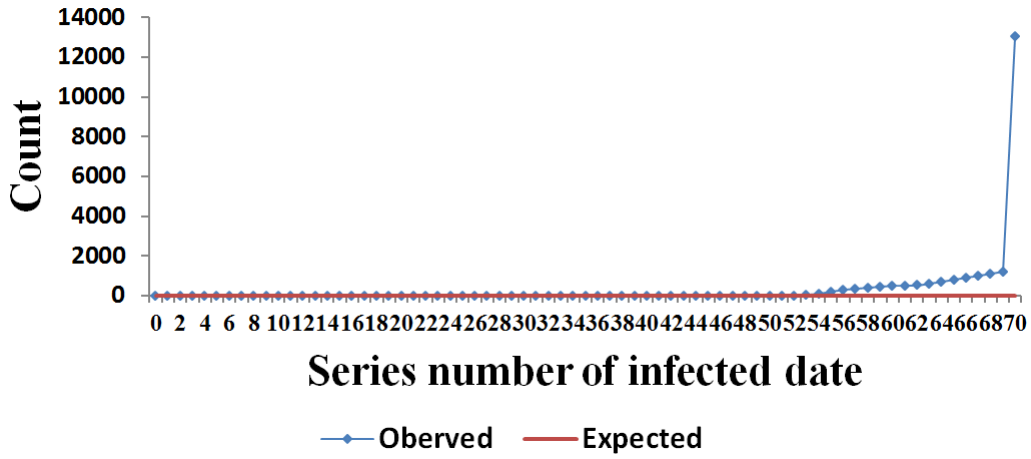


Figure 2 The acronym of COVID in the two models to display(EGV model(top) and IRT model(bottom))



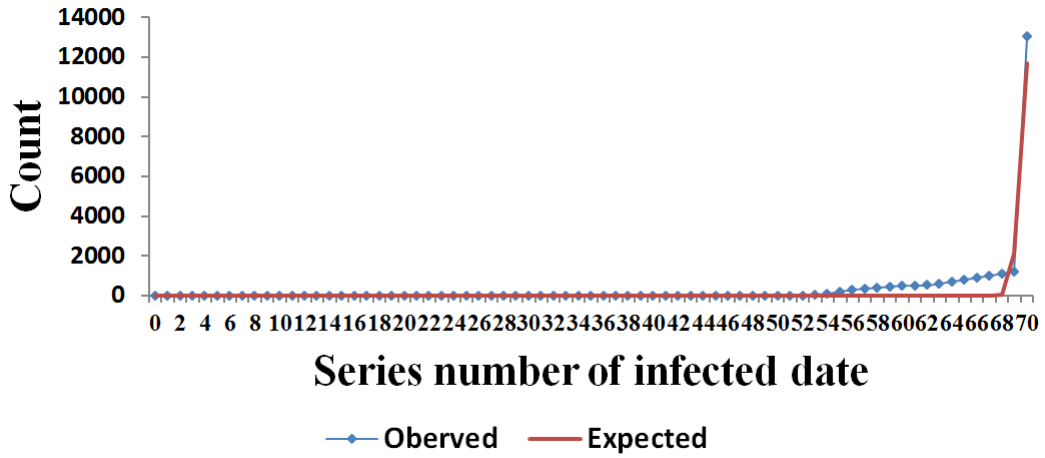


Figure 3 The acronym of COV in the two models to display(EGV model(top) and IRT model(bottom))

3.3 **The temporal bar graph to display the evolution of medical acronyms since 1950**

Using a temporal bar graph to display the evolution of the medical acronym is demonstrated in Figure 5. We can see that the black bubbles are the medians, indicating the most number of frequency observed in an acronym. Besides the top one with COVID as increasing owning to the median overlaid with the IP bubble in green, others are on the left side apart from the IP bubble, implying the decline in the number of acronyms. In contrast, if the black is on the right side of IP bubble, the phenomenon is to decrease or to the stationarity at the end. The inner bars show the number of the acronym in red color, the darker means the more number of acronym observed in the past. The cutting point is set at 330. If the number is less than 330, no inner bar appears on the light outer bar. Readers are invited to click on the link[11] to examine the details on Google Maps.

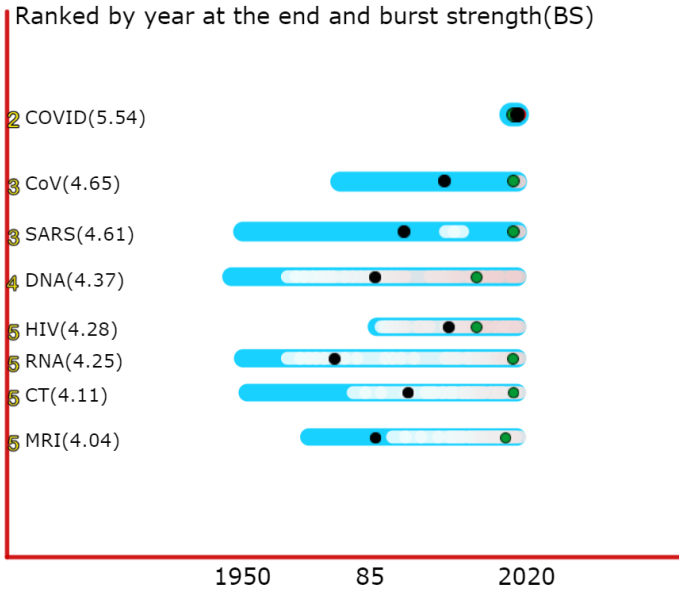


Figure 4 Using temporal bar graph to display the evolution of medical acronyms

Discussions and conclusion

The origianl derivation of the EGV model for the IP determination is verified in the reference[11]. The IP determination was present in the references[6,7] using the AAC[8], which is tedious and computation burden in searching IP on an ogive curve instead of the item difficulty with the direct approach and fast way to determine the IP as we did in this study. The comparison of R-square and IPs in the two mathematical models of IRT and EGV is novel and modern and provides deeper insights into the feature of IP determinations. The IRT model is superior to the EGV model in R-square and recommended to future research, not limited to the bibliometric analysis only, such as COVID-19 situation with IP to denote the containment of epidemics in countries/regions.

Reference

1. Barnett, A., Doubleday, Z. (2021). Demonstrating the ascendancy of COVID-19 research using acronyms. Scientometrics, 16, 1–4. 10.1007/s11192-021-04016-8.
2. Kuo SC, Chien TW, Chou W. Questions to the article: demonstrating the ascendancy of COVID-19 research using acronyms. Scientometrics. 2021 Aug 5:1-4. doi: 10.1007/s11192-021-04108-5. Epub ahead of print. PMID: 34376877
3. Shen L, Xiong B, Li W, Lan F, Evans R, Zhang W. Visualizing Collaboration Characteristics and Topic Burst on International Mobile Health Research: Bibliometric Analysis. JMIR Mhealth Uhealth. 2018 Jun 5;6(6):e135.
4. Kleinberg J. Bursty and hierarchical structure in streams. Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; KDD '02; July 23-26, 2002; Edmonton, AB, Canada. 2002. pp. 91–101. [CrossRef] [Google Scholar]
5. 38. Sci2 Team. 2009. [2017-05-27]. Sci2 tool : a tool for science of science research and practice https://sci2.cns.iu.edu/user/index.php webcite.
6. Lee KW, Chien TW, Yeh YT, Chou W, Wang HY. An online time-to-event dashboard comparing the effective control of COVID-19 among continents using the inflection point on an ogive curve: Observational study. Medicine (Baltimore). 2021 Mar 12;100(10):e24749.
7. Wang LY, Chien TW, Chou W. Using the IPcase Index with Inflection Points and the Corresponding Case Numbers to Identify the Impact Hit by COVID-19 in China: An Observation Study. Int J Environ Res Public Health. 2021 Feb 18;18(4):1994.
8. Yang DH, Chien TW, Yeh YT, Yang TY, Chou W, Lin JK. Using the absolute advantage coefficient (AAC) to measure the strength of damage hit by COVID-19 in India on a growth-share matrix. Eur J Med Res. 2021 Jun 24;26(1):61.
9. Wiki Definition of an Inflection Point. [(accessed on 22 October 2020)]; Available online: https://en.wikipedia.org/wiki/Inflection_point.
10. Chien TW. Using Solver in Microsoft Excel to estimate parame-ters in Rasch model: A technique study. J Bibliographical Analyses in Statistics 2021; 18(3): 5-8. 2021/8/24 at http://www.healthup.org.tw/article/articlex/raschsolvor.pdf
11. Chien TW. Abstract video in this study. 2021/8/14 at https://youtu.be/Cc4zTK7i7SI
12. Chien TW. Figure 4 in this study. 2021/8/14 at http://www.healthup.org.tw/gps/covidbargraphB2.htm
13. Shao Y, Chien TW. The determination of inflection curve on a given ogive curve using the second order derivative in calculus. J Bibliographical Analyses in Statistics 2021; 18(3): 21-23.