科技部補助專題研究計畫成果報告
期末報告

以信心強度為底之適測方法閾值估計探究

計 畫 類 別 ： 個別型計畫
計 畫 編 號 ： MOST 105-2410-H-002-056-
執 行 期 間 ： 105年08月01日至106年07月31日
執 行 單 位 ： 國立臺灣大學心理學系暨研究所

計畫主持人： 徐永豐

計畫參與人員： 大專生-兼任助理：鄭澈
其他-兼任助理：林易萱

報 告 附 件 ： 出席國際學術會議心得報告

中 華 民 國 106 年 10 月 31 日

中 文 摘 要 ： 我們從方法學角度探討以信心強度為底之閾值估計法則，此法則可套用於現有之（無母數、固定梯級）適測方法。我們藉電腦模擬有系統操弄變項，探討此法則閾值估計之收斂性。我們以75%閾值估計為例，說明不管在大樣本(=150)還是小樣本(=50)情況下，閾值估計在梯級很小時皆大體良好。特別是，在小樣本情況下，閾值估計是否良好端視是否起始值距目標值不要太遠。

中文關鍵詞： 心理物理學; 信心強度; 閾值估計; 適測方法

英 文 摘 要 ： In this project we proposed, mainly methodologically, an alternative method for threshold estimation that is response-confidence embedded and can be implemented into existing (non-parametric, fixed-step-size) up-down methods. We investigated via simulation the convergence pattern of threshold estimation. Using the 75% threshold estimation as an example, we showed that for both large (n=150) and small (n=50) sample sizes, the estimate is by and large good for small step sizes. In particular, for small sample size the convergence depends more on whether the initial value is not far away from the target.

英文關鍵詞： psychophysics; response confidence; threshold estimation; up-down methods

Final Report

1 Introduction

The present project is a follow-up of Hsu and Chin (2014), in which we investigated the extension of some of the (non-parametric, fixed-step-size) *up-down* (also known as *staircase*) methods for threshold estimation proposed by other researchers (noteably Kaernbach, 2001; Klein, 2001). We characterized their extension by a general algorithm having the features of 'unbiased' middle response category as well as symmetrical weights for response confidence. We pointed out that these two features put severe limitations to the algorithm, often leading to biased and inconsistent estimates of threshold. Hsu and Chin (2014) also mentioned that it is possible to bypass the core assumptions at issue in the algorithm when incorporating response confidence into the up-down methods. The methodology has never been scrutinized and is the focus of this project. In the following we give a brief background introduction.

1.1 The Up-down method: From binary to multiple response categories

To begin with, consider a two-alternative forced choice (2AFC) discriminative task in which two alternatives, a *referent* and a *comparison* stimulus, are presented successively in different intervals (or simultaneously in different locations) on each trial. To facilitate our discussion, let $\mathbf{X}_n$ be a random variable representing the (physical) strength of the stimulus on trial n. We define a random variable $\mathbf{Y}_n$ for the participant's response on trial n by

$$\mathbf{Y}_n = \begin{cases} 1 & \text{if } \mathbf{X}_n \text{ is judged as exceeding the referent,} \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

We sometimes call $\mathbf{Y}_n = 1$ a *positive* response and $\mathbf{Y}_n = 0$ a *negative* response.

The gist of the family of up-down methods concerns adjusting the comparison stimulus on the current trial based on the participant's response on the previous trial. A well-known example is the simple up-down method $\mathbf{X}_{n+1} = \mathbf{X}_n - \delta(2\mathbf{Y}_n - 1)$ (where δ is the step size) for estimating the 50% threshold (Dixon and Mood, 1948). Another example is the 'weighted up-down' (WUD) method proposed by Kaernbach (1991) for estimating any threshold quantiles. Specifically, for any response criterion π $(0 < \pi < 1)$,

$$\mathbf{X}_{n+1} = \mathbf{X}_n - \delta\left(\frac{\mathbf{Y}_n - \pi}{\pi}\right). \tag{2}$$

In words, there is a fixed ratio $\frac{\pi}{1-\pi}$ of the increasing (after a negative response) and decreasing (after a positive response) step size throughout the sequence. For example, for the 75% threshold, after a negative response the signal will go up one step on the next trial, but will only go down 1/3 steps after a positive response.

Later Kaernbach (2001) extended WUD to include a "don't know" response option, which was recaptured by Hsu and Chin (2014) by setting $\mathbf{Y}_n = 0.5$ in Equation 1 to represent the "don't know" response option. i.e.,

$$\mathbf{Y}_n = \begin{cases} 1 & \text{if } \mathbf{X}_n \text{ is judged as exceeding the referent,} \\ 0.5 & \text{if ``don't know'',} \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Klein (2001) extended the case of Kaernbach (2001) via two examples to include four response categories, namely the high- and low-confidence judgment for each of the positive and negative responses. Hsu and Chin (2014) extended Equation 3 to cover the cases in Klein (2001). Specifically, let

$$\mathbf{Y}_n = \begin{cases} 1 & \text{if `positive' with high confidence,} \\ \alpha_2 & \text{if `positive' with low confidence,} \\ \alpha_3 & \text{if `negative' with low confidence,} \\ 0 & \text{if `negative' with high confidence,} \end{cases} \tag{4}$$

in which $\alpha_4 = 0 \leq \alpha_3 \leq \alpha_2 \leq 1 = \alpha_1$ satisfying $(\alpha_2 + \alpha_3)/2 = 0.5$. It can be easily shown that the two extended cases of WUD proposed by Klein (2001) are special cases of this algorithm.

1.2 An algorithm and its limitations

One can continue the extension of Equation 4 to include more than four response categories. The algorithm is as follows. For m response categories with 1 representing 'certainly positive' and m representing 'certainly negative,' we set the corresponding $\mathbf{Y}_n$ to range from $1 = \alpha_1, \alpha_2, \ldots, \alpha_m = 0$ satisfying $\alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_m$. We further require the α_i values to be symmetric around 0.5. i.e., if m is even, then $(\alpha_k + \alpha_{m-k+1})/2 = 0.5$ for $k = 1, \ldots, \frac{m}{2}$. If m is odd, then $\alpha_{\frac{m+1}{2}} = 0.5$ and $(\alpha_k + \alpha_{m-k+1})/2 = 0.5$ for $k = 1, \ldots, \frac{m-1}{2}$. Moreover, for each response category k ($k = 1, \ldots, m$) there is a function $p_k \geq 0$, and $\sum_{k=1}^{m} p_k(x) = 1$ for all x. This algorithm also can be applied to other up-down methods to incorporate response confidence. See Hsu and Chin (2014) for details.

While the algorithm just mentioned is simple and can be easily implemented, there is an issue of unbiasedness of threshold estimation. Hsu and Chin (2014) discussed two main features of the algorithm that limit its usability. The first is concerned with the characteristic of having 0.5 for the middle response category, and the second is concerned with the choice of weights for response-confidence levels. Using a series of simulation based on the discrete-state threshold theory of Hsu and Doble (2015) for the underlying psychometric function, Hsu and Chin (2014) demonstrated that the algorithm indeed is somewhat biased for threshold estimation, especially if the weight of the middle response category is not set at 0.5 (or is not symmetric about 0.5). This statement also was supported by a small scale of experiment (see Hsu and Chin, 2014, for details). Since having biased levels of response confidence may be a norm but not an anomaly in real situations, it is very likely that the algorithm would yield somewhat biased estimates of threshold.

1.3 An alternative method

Hsu and Chin (2014) mentioned in their Concluding Remarks section that it is possible to embed response confidence in the up-down methods and at the same time bypass the the feature of symmetrical confidence levels in the algorithm. The idea is to let the experimenter adopt a fixed cutoff throughout the trials, and on each trial the participant simply responds to the stimulus by marking her response on a continuous scale for her confidence level. The experimenter then assigns the stimulus on the next trial based on a predetermined cutoff on the scale, which the participant is unaware of, together with any of the (original) binary-based algorithms of the up-down methods for a given response criterion.

Specifically, continuing the refinement of the number (m) of categories for response confidence, as $m \to \infty$ we have a continuous version of response confidence. Let the random variable $\mathbf{R}_n$ denote the response confidence on trial n, and let ν be a real number representing the fixed cutoff. We now define

$$\mathbf{Y}_n = \begin{cases} 1 & \text{if } \mathbf{R}_n > \nu, \\ 0 & \text{if } \mathbf{R}_n \leq \nu. \end{cases} \tag{5}$$

In words, the stimulus value that elicits a certain criterion (π) of response confidence at a given cutoff (ν) can be estimated. Doing this is tantamount to introducing another dimension to the existing framework. Equation 5 can be implemented to any of the (binary) up-down algorithms, such as the WUD (Kaernbach, 1991) and the 'biased coin design' (Durham and Flournoy, 1995). In fact, the up-down methods just mentioned can be regarded as a special case of Equation 5 in which the cutoff ν is set in the middle of the scale separating the positive and negative responses.

In this project we aim to study, mainly methodologically, the applicability of a response-confidence embedded method based on Equation 5, jointly with some of the up-down methods discussed in Hsu and Chin (2014), for threshold estimation.

2 Procedure

Our earlier study (Chen and Hsu, 2010) showed via simulation that the 'biased-coin design' (BCD), originally developed by Durham and Flournoy (1995) for experimenting drug dosages in clinical trials, performed well with small step size, even for small trial sequences. Thus in this project we mainly focus on applying Equation 5 to the BCD.[1] We also focus on the response criteria between 0.5 and 1, as it is often used in real situations. The algorithm of BCD is as follows. For $0.5 \leq \pi < 1$,

$$\mathbf{X}_{n+1} = \mathbf{X}_n - \delta\left(\mathbf{Y}_n(\mathbf{S}_\pi + 1) - 1\right), \tag{6}$$

where $\mathbf{S}_\pi$ is a Bernoulli random variable (independent of $\mathbf{Y}_n$) taking value 1 with probability $\frac{1-\pi}{\pi}$ and value 0 with probability $1-\frac{1-\pi}{\pi}$. In words, a positive response by the participant will either yield a one-step down (with probability $\frac{1-\pi}{\pi}$) or no change of step size (with probability $1-\frac{1-\pi}{\pi}$) for the signal on the next trial. In contrast, a negative response by the participant will yield a one-step up for the signal on the next trial.

Following our notation of representing 0 for negative response and 1 for positive response, we conveniently set the range of response confidence to be between 0 and 1, and used a truncated Gaussian (between 0 and 1) for each stimulus-elicited underlying distribution of response confidence. The standard deviation of the truncated Gaussian was set at 0.1. The (physical) signal intensity is a ratio

[1] We also applied the algorithm to the WUD and found that the convergence pattern is similar to that of BCD.

scale and here we conveniently assumed that a 0.5 units of signal intensity would elicit an average of 0.5 (i.e., the middle point) on the response confidence scale. Under this constraint and the observation that typical psychometric functions (of signal intensity) are largely linear in the mid range, we used a logistically distributed psychometric function for the mean response confidence. The location and scale parameters of the logistic distribution were set at 0.5 and 0.2, respectively.[2] For comparison, we also ran another series of simulation based on a 'rigid' (with slope 1) psychometric function.

We systematically ran the simulation with different response criteria (π) and cutoffs (ν) for the BCD. We also paid special attention to the step size (δ) implemented in the up-down method, as previous studies have shown that different step sizes would affect the speed of convergence. In particular, large step sizes tend to yield biased estimates (e.g., Hsu and Chen, 2009; Hsu and Chin, 2014). After some exploration, we chose five step sizes: 0.04, 0.08, 0.12, 0.16, and 0.20 units in the simulation. All the simulation was performed using the R software (R Development Core Team, 2015).

3 Results and concluding remarks

A series of simulation was conducted to get an insight into the convergence pattern. For each condition of the simulation, we recorded the reversal points except the first two, and computed the average. When there was no more than one reversal point, we used the signal intensity in the last trial as the estimate. We then repeated the simulation procedure 1000 times and recorded the median.

We obtained many simulation results. To get straight to the point, we show in Figure 1 the estimation of the 75% threshold.[3] We first note that for large sample size (=150), the estimate is excellent for small step size, no matter the initial values. For small sample size (=50) the convergence depends on the initial value, and this applies to all the cutoffs (0.2, 0.5, and 0.8) we assigned. In other words, as long as the initial value is not far away from the target, the estimate is by and large a good approximation to the target for small step size. The simulation using the 'rigid' (with slope 1) psychometric function yielded slightly better results. See Figure 2 for details.

[2] More can be elaborated about the parameter setting of the truncated Gamma distributed response confidence, and also of the logistic distributed psychometric function (of signal intensity) for the mean response confidence. We do not go to the detail here.

[3] The simulation for other non-extreme percentile thresholds yielded similar results.

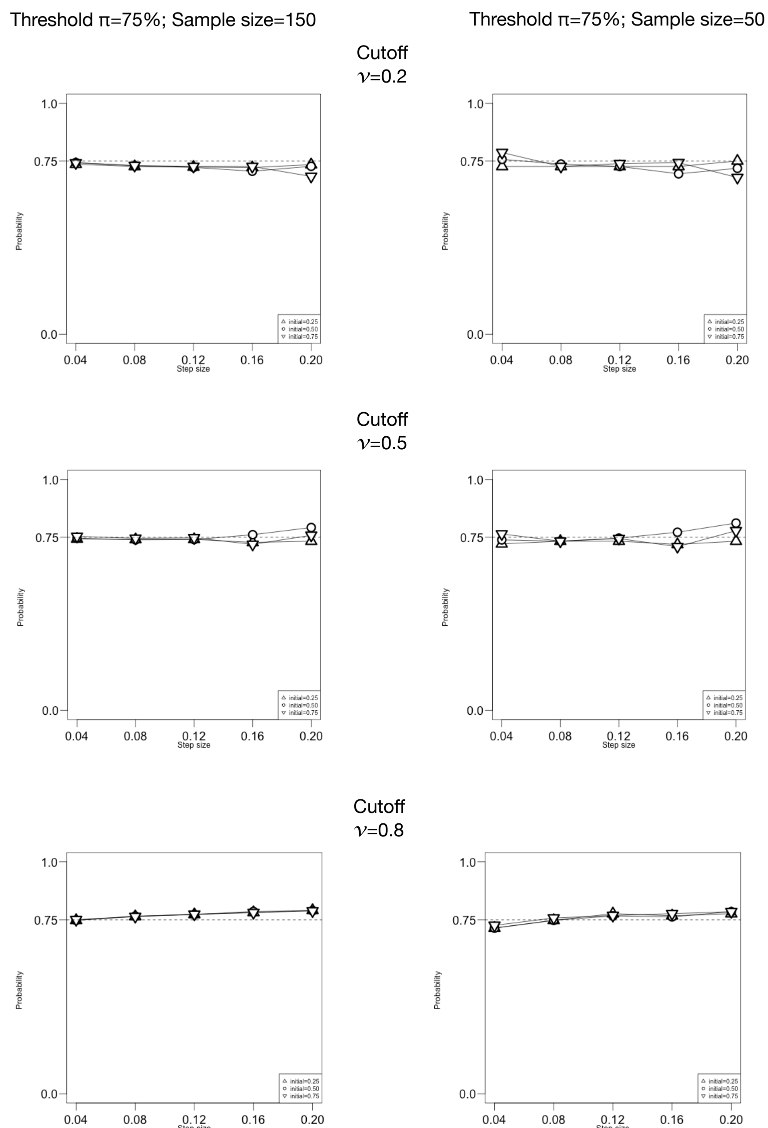

Figure 1: Estimates of the 75% threshold assuming a logistic psychometric function. Left column: sample size=150, right column: sample size=50.

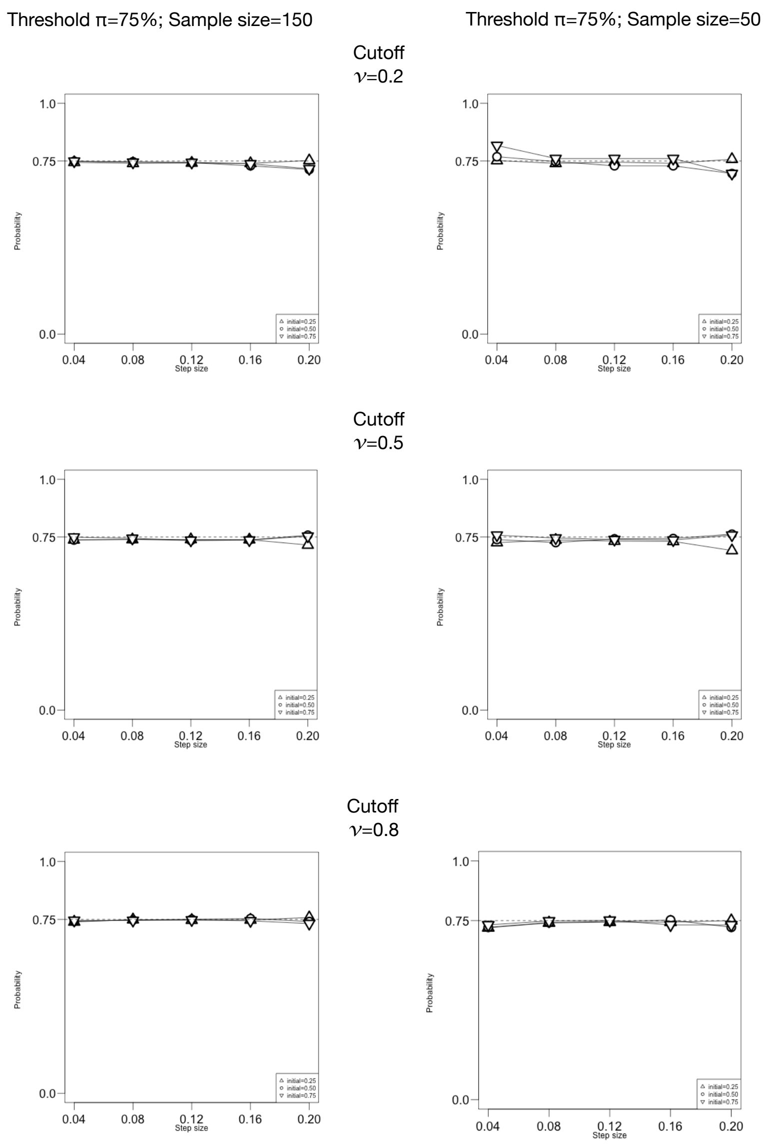

Figure 2: Estimates of the 75% threshold assuming a 'rigid' psychometric function (with slope 1). Left column: sample size=150, right column: sample size=50.

In this project we provided a alternative method for response-confidence embedded threshold estimation. The core of the method relies on putting response confidence into another dimension so that it would not confound with possible biasedness under the conventional up-down methods. We showed by simulation that the new method is potentially applicable. A future work is to implement the method to psychophysical tasks to verify its feasibility. The methodology will be useful in, for example, some clinical settings where participants from certain clinical groups might only contribute a small amount of trials.

References

Chen, Y.-H. and Hsu, Y.-F. (2010). Convergence of non-parametric adaptive threshold estimation in the yes-no task. *Chinese Journal of Psychology*, 52:95–112.

Dixon, W. and Mood, A. (1948). A method for obtaining and analyzing sensitivity data. *Journal of American Statistical Association*, 43:109–126.

Durham, S. D. and Flournoy, N. (1995). Up-and-down designs I: stationary treatment distributions. In Flournoy, N. and Rosenberger, W. F., editors, *Adaptive Designs*, pages 139–157. IMS Lecture Notes, 25. Hayward CA.

Hsu, Y.-F. and Chen, Y.-H. (2009). Applications of nonparametric adaptive methods for simple reaction time experiments. *Attention, Perception, & Psychophysics*, 71:1664–1675.

Hsu, Y.-F. and Chin, C.-L. (2014). On the limitations of fixed-step-size adaptive methods with response confidence. *British Journal of Mathematical and Statistical Psychology*, 67:266–283.

Hsu, Y.-F. and Doble, C. W. (2015). A threshold theory account of psychometric functions with response confidence under the balance condition. *British Journal of Mathematical and Statistical Psychology*, 68:158–177.

Kaernbach, C. (1991). Simple adaptive testing with the weighted up-down method. *Perception & Psychophysics*, 49:227–229.

Kaernbach, C. (2001). Adaptive threshold estimation with unforced-choice tasks. *Perception & Psychophysics*, 63:1377–1388.

Klein, S. (2001). Measuring, estimating, and understanding the psychometric function: A commentary. *Perception & Psychophysics*, 63:1421–1455.

R Development Core Team (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

出國報告（出國類別：出席會議）

出席 Psychonomic Society 年會

服務機關：國立台灣大學心理系
姓名職稱：徐永豐 副教授
派赴國家：Boston, USA
出國期間：November 17–20, 2016
報告日期：November 30, 2016
經費來源：科技部

目的

The main purpose of this trip was to present my work in the 57th Annual Meeting of the Psychonomic Society, which began the afternoon of November 17th and ended the early afternoon of November 20th in Boston. After the conference I flew to Irvine in Southern California for a research stay for several days.

過程

In the Psychonomic conference I gave a poster presentation titled "A response-confidence embedded method for threshold estimation". Based on some of my previous work (Hsu & Chen, 2009; Hsu & Chin, 2014), I proposed a response-confidence embedded method for threshold estimation in psychophysics. The proposed method can be incorporated into the family of non-parametric, fixed-step-size up-down algorithms (e.g., Derman, 1957; Durham & Flournoy, 1995; Kaernbach, 1991, 2001; Klein, 2001) for the estimate of threshold quantiles with certain levels of response confidence (set by some predetermined criteria).

心得

Attending the Psychonomic conference helps me keep track of the on-going research topics in the respective fields. For example, one of the symposia, "Introduction to Model-Based Cognitive Neuroscience" provided a glimpse of the current attempt of modeling behavioral and neuroimaging data jointly. And it appears the advancement would benefit from incorporating the Bayesian approach.

After the conference I flew to Southern California to meet with Dr. Chris Doble to discuss a potential joint project of studying the constraints of the law of similarity under the Fechnerian framework. This research initiative will be a follow-up of my recent work of characterizing the conditions under which affine representations are also Fechnerian under the power law of similarity (Hsu & Iverson, 2016).

105年度專題研究計畫成果彙整表

計畫主持人：徐永豐					計畫編號：105-2410-H-002-056-		
計畫名稱：以信心強度為底之適測方法閾值估計探究							
成果項目					量化	單位	質化（說明：各成果項目請附佐證資料或細項說明，如期刊名稱、年份、卷期、起訖頁數、證號...等）
國內	學術性論文	期刊論文			0	篇	
		研討會論文			0		
		專書			0	本	
		專書論文			0	章	
		技術報告			0	篇	
		其他			0	篇	
	智慧財產權及成果	專利權	發明專利	申請中	0	件	
				已獲得	0		
			新型/設計專利		0		
		商標權			0		
		營業秘密			0		
		積體電路電路布局權			0		
		著作權			0		
		品種權			0		
		其他			0		
	技術移轉	件數			0	件	
		收入			0	千元	
國外	學術性論文	期刊論文			0	篇	
		研討會論文			1		N/A
		專書			0	本	
		專書論文			0	章	
		技術報告			0	篇	
		其他			0	篇	
	智慧財產權及成果	專利權	發明專利	申請中	0	件	
				已獲得	0		
			新型/設計專利		0		
		商標權			0		
		營業秘密			0		
		積體電路電路布局權			0		
		著作權			0		
		品種權			0		
		其他			0		

<table>
<tr><td></td><td rowspan="2">技術移轉</td><td>件數</td><td>0</td><td>件</td><td></td></tr>
<tr><td></td><td>收入</td><td>0</td><td>千元</td><td></td></tr>
<tr><td rowspan="10">參與計畫人力</td><td rowspan="5">本國籍</td><td>大專生</td><td>1</td><td rowspan="10">人次</td><td>N/A</td></tr>
<tr><td>碩士生</td><td>0</td><td></td></tr>
<tr><td>博士生</td><td>0</td><td></td></tr>
<tr><td>博士後研究員</td><td>0</td><td></td></tr>
<tr><td>專任助理</td><td>0</td><td></td></tr>
<tr><td rowspan="5">非本國籍</td><td>大專生</td><td>0</td><td></td></tr>
<tr><td>碩士生</td><td>0</td><td></td></tr>
<tr><td>博士生</td><td>0</td><td></td></tr>
<tr><td>博士後研究員</td><td>0</td><td></td></tr>
<tr><td>專任助理</td><td>0</td><td></td></tr>
<tr><td colspan="4">其他成果
（無法以量化表達之成果如辦理學術活動、獲得獎項、重要國際合作、研究成果國際影響力及其他協助產業技術發展之具體效益事項等，請以文字敘述填列。）</td><td colspan="2"></td></tr>
</table>

科技部補助專題研究計畫成果自評表

請就研究內容與原計畫相符程度、達成預期目標情況、研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）、是否適合在學術期刊發表或申請專利、主要發現（簡要敘述成果是否具有政策應用參考價值及具影響公共利益之重大發現）或其他有關價值等，作一綜合評估。

1. 請就研究內容與原計畫相符程度、達成預期目標情況作一綜合評估
   ■達成目標
   □未達成目標（請說明，以100字為限）
      □實驗失敗
      □因故實驗中斷
      □其他原因
   說明：

2. 研究成果在學術期刊發表或申請專利等情形（請於其他欄註明專利及技轉之證號、合約、申請及洽談等詳細資訊）
   論文：□已發表 □未發表之文稿 ■撰寫中 □無
   專利：□已獲得 □申請中 ■無
   技轉：□已技轉 □洽談中 ■無
   其他：（以200字為限）

3. 請依學術成就、技術創新、社會影響等方面，評估研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性，以500字為限）
   The response-confidence embedded up-down method we developed (for threshold estimation) will be useful in, for example, some clinical settings where participants from certain clinical groups might only contribute a small amount of trials.

4. 主要發現
   本研究具有政策應用參考價值：■否 □是，建議提供機關
   （勾選「是」者，請列舉建議可提供施政參考之業務主管機關）
   本研究具影響公共利益之重大發現：□否 □是
   說明：（以150字為限）