

國立臺灣大學理學院心理學研究所



碩士論文

Graduate Institute of Psychology

College of Science

National Taiwan University

Master Thesis

回答信心量尺與適測方法結合於閾值估計實驗之影響

The Influence of Combination of Response Confidence Scale with
Adaptive Methods in Threshold Estimation Experiments

黃妍君

Yan-Jun Huang

指導教授：徐永豐 博士

Advisor: Yung-Fong Hsu, Ph.D.

中華民國 108 年 9 月

September 2019



致謝

感謝這三年多的時間裡，許多人的幫助和徐老師的盡心指導，實驗室同學的熱心幫助，以及我身邊朋友們的溫馨支持。我在這裡學習到了很多心理學知識，認識了許多人，也遇見了很多有趣的事情，這些都將成為我人生裡溫暖的回憶。

我要感謝徐老師一直耐心的給予我指導，讓我能一步一步的認識到心理物理學的有趣之處。同時也感謝實驗室的朋友們的幫助，在一次次的報告裡給予我寶貴的建議。因為這些幫助我才能如願完成我的論文；不僅如此，我在和大家的交流裡也學習到了很多有趣的新知識，也了解了許多看問題的新角度。我也很高興自己選擇心理學，這門學科開闊了我的視野，也帶給我很多不同的切入問題的角度。

最後，我要感謝那些我在這裡結識的許多人，也感謝在遠方的家人、朋友所給予我的支持和溫暖。



摘要

在一個閾值實驗裡閾值通常以適測方法估計；這些實驗是基於二選項回答（「是」/「否」），但二選項回答所帶來的一些限制會影響實驗結果。為解決此問題，學者試圖通過擴展回答選項個數的方式將回答信心包含進實驗裡（Hsu & Chin, 2014; Kaernbach, 2001; Klein, 2001）。跟隨該研究方向，在這篇論文裡，我採用電腦模擬探討在兩種實驗（「是」/「否」作業與「同」/「異」作業）情境中引入一個連續回答信心量尺對不同適測方法（適測方法包含帶權重上下法、偏差硬幣投擲法與德爾曼上下法）的影響。更進一步地，我經由系統性地操作初始值、步長、試驗數等實驗變量探討它們對適測方法的影響。模擬結果表明帶權重上下法整體比另兩個適測方法表現好。值得注意的是，回答信心量尺的截點位置會影響適測方法的表現，且在一些條件下會引發偏誤的出現。我們研究表明通過調整實驗變量可以減小甚至消除這些偏誤。綜上，在閾值實驗裡引入回答信心量尺可給予研究者在實驗設計階段更多的自由，其附帶偏誤亦存在合理解決方式以減小其影響。

關鍵詞：適測方法、閾值估計、回答信心、帶權重上下法、偏差硬幣投擲法、德爾曼上下法

The Influence Of Combination Of Response Confidence Scale With Adaptive Methods In Threshold Estimation Experiments

Yan-Jun Huang



Abstract

In a threshold estimation experiment, the threshold is commonly estimated by adaptive methods. These experiments are based on binary responses (Yes/No). However, there are some limitations of binary responses that could affect the results of experiments. To solve this problem, scholars have tried to include the role of response confidence in the experiment by extending the number of category of responses (Hsu & Chin, 2014; Kaernbach, 2001; Klein, 2001). Following this idea, in this thesis I used simulation to survey the influence of introducing a continuous response confidence scale to different adaptive methods (including the Weighted Up-Down method, Biased Coin Design, and Derman's Up-Down method) in two types of threshold experiments (Yes/No tasks and Same/Difference tasks). Moreover, I systematically manipulated the initial value, step size, trial number and other experiment variables to discuss their influences on adaptive methods. The simulation results show that the performance of the Weighted Up-Down method is better than the other two adaptive methods. Note that the placement of cut-offs on the response confidence scale might affect the performance of adaptive methods, and in some cases leads to the emergence of biases. We show that these biases can be reduced or even eliminated by adjusting some experiment variables. To summarize, the introduction of the response confidence scale can provide researchers with more freedom on experimental design, and there exists reasonable solutions to reduce the potential biases generated by this scale.

Keywords: adaptive methods, threshold estimation, response confidence, Weighted Up-Down method, Biased Coin Design, Derman's Up-Down method



目次

第一章 緒論	1
第二章 文獻回顧	5
第一節 閾值估計實驗的「是」/「否」作業與「同」/「異」作業	5
第二節 適測方法	6
第三節 回答信心	13
第三章 模擬實驗	17
第一節 回答信心的構建	17
第二節 3D/2D繪圖	20
第三節 適測方法、截點、目標概率、步長、初始值、試驗數、模擬次數	25
第四節 均方根誤差與偏誤	27
第四章 模擬結果	29
第一節 估計目標概率圖	29
第二節 均方根誤差	33
第三節 偏誤	37
第五章 擴展	42
第一節 「同」/「異」作業與其非單調心理計量函數	42
第二節 非單調心理計量函數中適測方法的應用	44
第三節 適測方法、截點、目標概率、步長、初始值、試驗數、模擬次數	50
第四節 結果	50
第六章 總結	62
第七章 參考文獻	65
第八章 附錄	67
附錄一 「是」/「否」作業下均方根誤差與偏誤的最大值/最小值	67
附錄二 「同」/「異」作業下均方根誤差與偏誤的最大值/最小值	68
附錄三 Hsu與Doble (2015) 方法在「同」/「異」作業的應用	68



表圖目次

表 1 「是」/「否」作業下均方根誤差與偏誤的最大值/最小值	67
表 2 「同」/「異」作業下均方根誤差與偏誤的最大值/最小值	68
圖 1 3D模型截圖	23
圖 2 2D分圖-回答信心分布函數的概率密度函數	24
圖 3 2D分圖-回答信心的期望值與刺激強度的關係	24
圖 4 2D分圖-心理計量函數與其對應的截點關係	25
圖 5 偏差硬幣投擲法-估計目標概率圖	32
圖 6 德爾曼上下法-估計目標概率圖	34
圖 7 帶權重上下法-估計目標概率圖	35
圖 8 偏差硬幣投擲法-均方根誤差	39
圖 9 德爾曼上下法-均方根誤差	40
圖 10 帶權重上下法-偏誤	41
圖 11 「同」/「異」作業裡的心理計量函數	43
圖 12 偏差硬幣投擲法-估計目標概率圖	55
圖 13 德爾曼上下法與帶權重上下法-估計目標概率圖	56
圖 14 偏差硬幣投擲法-均方根誤差	58
圖 15 德爾曼上下法-均方根誤差	59
圖 16 帶權重上下法-偏誤	60
圖 17 偏差硬幣投擲法、德爾曼上下法-偏誤	61



第一章 緒論

心理物理學研究的是人的心理感受與實際物理刺激強度之間的關係，描述其關係的函數即心理計量函數（psychometric function）。繪製完整的心理計量函數或僅估計該函數的某個數據點可通過「是」/「否」作業（Yes/No task）完成。常見的「是」/「否」作業的實驗方法具體如下：在每次試驗裡，研究者向受試者呈現兩個物理刺激，一個刺激在試驗中會不斷發生變化，另一刺激則恆定不變；並且兩個刺激的左右位置隨機分配。受試者必須在每次刺激呈現結束後，以兩個回答選項（「是」或「否」）的方式回答左邊的刺激是否比右邊的刺激強度大，其選擇結果被記錄再呈現下一組刺激。考慮到實際研究裡需估計心理計量函數的某個數據點的情況比需估計整條心理計量函數的情況更常見，本文僅探討以獲取心理計量函數中某一預先指定的數據點所對應的物理刺激強度為實驗目的的實驗，由於該數據點對應的物理刺激強度稱為「閾值」（threshold），此類型實驗稱為「閾值估計」（threshold estimation）實驗。

在閾值估計實驗裡受試者只能給予「是」或「否」的回答，這種二選項回答模式使其在實驗中有所侷限。在此將從兩種情境說明其侷限性（以下假定兩種情境裡左邊是變化刺激，右邊是恆定刺激。該設定不會影響情境說明，僅方便表述。）：首先，實驗要求受試者在每次試驗必須給出回答，當變化刺激與恆定刺激的差值很小時，受試者難以判定真實情況（即，左邊的刺激是否比右邊的刺激強度大），他極其可能隨機猜測給予回答。反之，當兩種刺激的差值較大時，受試者對自己給出的回答較為肯定，即受試者對自己的回答的信心程度很高。在上述兩種情境裡，受試者都可能給出回答「是」的結果，但受試者給出肯定回答時

的信心程度並不一致。

然而，由於研究者僅記錄回答選項（「是」或「否」），這兩種不同的情境最終會被歸類為同一種情境（即給予肯定回答的情境）。不僅如此，每位受試者對於兩個刺激差值的敏感程度並不相同，也使得研究者難以判斷受試者對回答結果的確定程度。就研究目的而言，為保證實驗效度，研究者需確保受試者給予的回答盡可能準確且具有一致性，因此上述兩種情境最好分開討論。

以解決此問題為目標，學界從回答選項的角度切入試圖分開這兩種情境：Kaernbach（2001）提出在原先兩個回答選項（「是」或「否」）的基礎上增加新的第三回答選項「不知道」，意即當受試者無法判斷兩個刺激的強度差異時，可選擇「不知道」避免猜測作答。Klein（2001）則將Kaernbach（2001）原有的三個回答選項擴展為四個回答選項，將「不知道」拆分成「低信心的是」與「低信心的否」以更加細分受試者在低信心度的情境。同時，Klein（2001）也指出三回答選項可能會引入回答偏誤，此處的回答偏誤指即便在信心程度僅是略低而不是特別低時，受試者仍更傾向給予「不知道」的回答選項而不是其他兩個回答選項。而根據受試者得到的實驗指導，此時受試者僅稍有不確定，他或許不應回答「不知道」。

儘管上述方法合理地分離受試者信心程度高低的情境，但在方法學的應用上仍需作調整。為具體說明此情況，在此先簡述閾值估計實驗裡適測方法的應用。在實驗情境裡，當受試者對該次試驗給予回答後，研究者根據受試者的回答決定下次試驗呈現的刺激強度，這種依賴於受試者的回答反饋決定下一次呈現的刺激強度的方法是「適測方法」（adaptive methods）。在適測方法中，研究者將受試者的回答帶入計算刺激強度的迭代式中得到下次試驗的刺激強度呈現給受試者；且現有的適測方法多適用於兩個回答選項。也就是說，若採用Kaernbach（2001）的三回答選項或Klein（2001）的四回答選項，原本以二回答選項為基礎的適測方法均需修改。

考慮到新增回答選項需修改適測方法的困擾，Hsu與Chin（2014）提出另一種分離高低信心程度的方法可避免對適測方法進行修改。由於引入附加回答選項的目的是更詳細地蒐集受試者回答的信心程度，他們提出若將回答選項數從多個擴展至無限，它可視作一個連續的回答量尺（以下稱為回答信心量尺）。研究者會預先在該量尺上設定一個固定點，本文定義為「截點」（cut-off）且受試者並不知道該截點存在。受試者在該回答信心量尺上標定自己主觀認為「左邊的刺激是否比右邊的刺激強度大」的信心程度。當受試者標定的信心程度超過這個截點時，研究者將該回答認定為等同於兩個選項回答當中的肯定回答「是」；反之就認定為「否」。因此，儘管受試者在連續的回答信心量尺上作答，最後在記錄數據時依舊保持二回答選項模式（受試者的回答僅被標記為「是」或「否」）。這種記錄方式使它能適用於所有基於二選項回答的適測方法且無需更改適測方法本身。

由於目前尚未有較為完整的研究將Hsu與Chin（2014）的想法應用於多種適測方法，所以該回答信心量尺對不同適測方法的影響尚不明確。因此，本文將延續Hsu與Chin（2014）的方式，選取德爾曼上下法（Derman's Up-Down method, Derman, 1957）、偏差硬幣投擲法（Biased Coin Design, Durham & Flournoy, 1995）和帶權重的上下法（Weighted Up-Down method, Kaernbach, 1991）三類適測方法。本文將在文獻回顧詳述此三類適測方法的流程及選取它們的理由。

本文的研究旨在採用模擬實驗來探討引入該回答信心量尺對於這些適測方法在閾值估計中的影響，同時系統性操作其它相關因素：步長（step size），初始值（initial value）和試驗數（trial number）以觀察這些因素在其中的作用。因此。本文在文獻回顧部分會先簡要陳述實驗作業的相關背景，並說明三種適測方法的流程，最後回顧以往學者對多種回答選項的相關研究。在模擬實驗詳述該量尺與心理計量函數間的關係以及模擬實驗的細節設定，最後加以分析模擬結果。

此外，該回答信心量尺不僅可用於「是」/「否」作業中，也可適用於

「同」/「異」作業 (Same/Difference task)。「同」/「異」作業是在相同的實驗背景下，受試者使用兩個回答選項（「是」或「否」）的方式回答變化刺激是否同恆定刺激的強度一樣。因此考慮到對該量尺應用性的豐富，本文會在擴展部分探討在「同」/「異」作業中引入該量尺對上述三種適測方法的影響。此前García-Pérez (2014) 提出另一種適用於「同」/「異」作業的適測方法，本文將在擴展部分將其與引入回答信心量尺的其他適測方法作比較並探討各自的偏差側重與優點不足。



第二章 文獻回顧

第一節 閾值估計實驗的「是」/「否」作業與「同」/「異」作業

傳統的心理物理學實驗可根據實驗設計的方式分為「是」/「否」作業與「同」/「異」作業。現假設以下實驗情境，研究者在每次試驗中呈現給受試者兩個刺激：其中之一在每次試驗中保持不變，另一刺激的強度則不斷變化；保持不變的刺激稱為「標準刺激」（referent stimuli），變化的刺激稱為「比較刺激」（compared stimuli）。並且每次試驗中這兩個刺激的位置會隨機分配。在「是」/「否」作業裡，受試者在每次試驗結束時需用「是」或「否」回答問題「左邊的刺激是否比右邊的刺激強度大」。在「同」/「異」作業裡，受試者在每次試驗結束時需用「是」或「否」回答問題「左邊的刺激是否與右邊的刺激強度一樣」。可以發現，兩種作業均以二回答選項的模式，即受試者僅以「是」或「否」回答研究問題，這也是此類實驗中常用的回答模式。

另一方面，從實驗目的來看，上述兩類作業均可用於以閾值估計為目的的心理物理學實驗。閾值估計實驗的心理計量函數通常以比較刺激與標準刺激的強度差為自變量，以受試者回答「是」的概率為因變量。本文「緒論」已前述閾值估計實驗針對心理計量函數裡預先決定的某一數據點，該數據點對應的刺激強度即為「閾值」，而該數據點對應的心理計量函數值即為「目標概率」。研究者在閾值估計實驗中預先確定目標概率，研究目的是找到該目標概率在心理計量函數上對應的刺激強度，即「閾值」。

在閾值估計的實驗裡，不同的實驗設計存在與其對應的心理計量函數。因

此，「是」/「否」作業與「同」/「異」作業的實驗設計差異導致這兩種作業對應的心理計量函數並不相同。具體而言，「是」/「否」作業的心理計量函數通常滿足單調性：從比較刺激強度從遠小於標準刺激，到比較刺激強度遠大於標準刺激的過程中，受試者主觀認為比較刺激比標準刺激強度大的概率也隨之增加。

需在此說明，因為在實驗中比較刺激與標準刺激的位置並不固定，並且受試者僅判斷左邊刺激是否比右邊刺激強度大。但研究者記錄的刺激強度是比較刺激與標準刺激的差值（即刺激強度=比較刺激-標準刺激），因此在實驗分析的階段會對受試者的回答稍作處理。具體而言，當比較刺激出現在右邊且標準刺激出現在左邊時，該試驗的刺激強度是比較刺激與標準刺激的差值，但受試者回答的「是」與「否」會被反向記錄：即「是」被記錄為「否」，「否」被記錄為「是」。因此，刺激強度從負向到正向的變化過程中，受試者被記錄回答「是」的概率會相應增加。

「同」/「異」作業裡，在同樣的刺激變化過程中，受試者認為比較刺激與標準刺激強度一樣大的概率隨比較刺激與標準刺激的強度逐漸接近而增大，而後受試者認為比較刺激與標準刺激強度一樣大的概率又因比較刺激超過標準刺激越來越多而減小。因此，「同」/「異」作業的心理計量函數並不具有單調性。儘管在此看來心理計量函數是否滿足單調性並不會對實驗進程造成任何影響，卻會影響實驗中適測方法的應用，此問題在接下來「適測方法」部分詳述。

第二節 適測方法

適測方法的主要目的是減少閾值估計實驗的試驗數，以提高實驗效率並節省資源。多數現有的適測方法要求心理計量函數是單調遞增函數且適用於二選項回答模式。Klein (2001) 將適測方法分為兩種：一種以「最

大概似估計」 (maximum likelihood) 為基礎的適測方法，另一種以「上下階梯法」 (Up-Down staircase) 為基礎的適測方法。其中，上下階梯法最先由Dixon與Mood (1948) 提出，它通過如下迭代式 (等式(1)) 計算下一次試驗的刺激強度。

$$X_{n+1} = X_n - \delta(2Y_n - 1) \quad (1)$$

其中， X_n, X_{n+1} 分別代表當前試驗的刺激強度差，下一次試驗的刺激強度差 (刺激強度差指的是上述實驗中比較刺激與標準刺激間的差值)； Y_n 的取值由受試者回答決定：在二回答選項的「是」/「否」作業中，若受試者回答「是」，則取值1，否則為0。 δ 是適測方法的步長。

由迭代式 (等式(1)) 可看出，上下階梯法的步長是一個恆定值，在整個實驗中保持不變。事實上這不是一個必須要求，適測方法的步長可根據當前已進行的試驗數進行調整，但本文中探討的三種適測方法均以上下階梯法為基礎且步長固定不變。這種選擇主要是考慮到如下三點：其一，減少適測方法間的根本性差異對結果的影響；其二，由於非固定步長的適測方法的步長會隨試驗數上升而減小，若該遞減速率過快，在試驗數過大的實驗裡則可能會出現收斂過慢的情況。其三，與非固定步長的適測方法相比，固定步長的適測方法往往會決定最終的實驗表現。Chen與Hsu (2010) 通過將非固定步長的快速隨機接近法 (Accelerated Stochastic Approximation, 即ASA) 與固定步長的偏差硬幣投擲法結合的方式阻止步長無限減小：它先設定一個較大的步長作為快速隨機接近法的起始步長，使其快速接近閾值所在的刺激強度附近，經過一定的試驗數，待步長下降到預先設定的標準後，切換使用偏差硬幣投擲法以保持固定步長，維持收斂速度。儘管這兩種方法的結合表現很好，但快速隨機接近法可以和所有的固定步長的適測方法相結合。因此，這種結合的最終結果取決於單個固定步長的適測方法本身。從以上角度考慮，本文只討論固定步長的適測方法。

與Klein (2001) 不同，Tretwein (1995) 更早提出另一種分類適測方法的標準，他從心理計量函數的性質入手，根據是否預先假設心理計量函數的函數形態分為母數適測方法和無母數適測方法。母數適測方法要求假設心理計量函數的函數形態，但無母數適測方法對心理計量函數的函數形態無前提假設。本文的研究因引入連續的回答信心量尺探討其在適測方法中的影響，此操作改變了實驗設計，導致可能會改變心理計量函數的形態，因此本文只考慮無母數適測方法。綜合以上兩點，本文選取的三種適測方法均為固定步長的無母數適測方法。

此外，本文中採用的三種固定步長的無母數適測方法在回答模式上均有相同的前提要求，即此三類適測方法均用於二選項回答模式。此三種適測方法亦要求滿足心理計量函數為遞增函數的條件，但「同」/「異」作業的設計模式使其心理計量函數不再滿足單調性，因此若在此類作業中使用傳統的適測方法需稍作調整以滿足其合理性。

針對「同」/「異」作業的心理計量函數無法滿足適測方法前提假設的問題，現階段可以通過調整心理計量函數或修改上下階梯法兩種方式解決。Hsu與Doble (2015) 從心理計量函數的角度入手，他們將非單調的心理計量函數轉化為單調遞增的心理計量函數，由此滿足傳統適測方法的前提條件以避免修改適測方法本身。García-Pérez (2014) 反向考慮此問題，他在以上下階梯法為核心的基礎上提出新的適測方法以避免對心理計量函數本身的轉換。這兩類做法的細節會在擴展部分詳述。

最後，此三類適測方法的差異主要體現在迭代式的差異，即計算下一次試驗的刺激強度的迭代式不同。以下將分別詳述此三類適測方法，即德爾曼上下法、偏差硬幣投擲法和帶權重上下法。



壹、德爾曼上下法

在「是」/「否」作業背景下，德爾曼 (Derman, 1957) 對前述的上下階梯法進行擴展，提出用於估計回答變量的累積分布函數上任意分位數對應的刺激值的德爾曼上下法。其中，該累積分布函數即心理計量函數，其分位數即目標概率，對應的刺激值即閾值。當選定的目標概率 (π) 不低於0.5時，該方法採用以下迭代式計算下一次試驗的刺激強度：

$$X_{n+1} = X_n - \delta(2Y_n S_\pi - 1) \quad (2)$$

其中， δ 是步長； S_π 是一個伯努利隨機變量，它有 $\frac{1}{2\pi}$ 的概率取值為1，有 $1 - \frac{1}{2\pi}$ 的概率取值為0。 Y_n 是一個回答變量，當受試者給予肯定回答時取值為1，反之取值為0。從迭代式中可以得出：當受試者給予肯定回答時，下一次試驗的刺激強度有 $\frac{1}{2\pi}$ 的概率會降低一個步長的值；有 $1 - \frac{1}{2\pi}$ 的概率會增加一個步長的值。當受試者給予否定回答時，下一次試驗的刺激強度一定增加一個步長的值（本文以下迭代式具有相同的數學符號均代表與此相同的數學意義）。

而當選定的目標概率低於0.5時，該方法採用以下迭代式計算下一次試驗的刺激強度：

$$X_{n+1} = X_n - \delta(1 - 2(1 - Y_n)S_{1-\pi}) \quad (3)$$

其中， S_π 是一個伯努利隨機變量，它有 $\frac{1}{2(1-\pi)}$ 的概率取值為1，有 $1 - \frac{1}{2(1-\pi)}$ 的概率取值為0；其他則與上述一致。由此表達式可以得出：當受試者給予肯定回答時，下一次試驗的刺激強度一定降低一個步長的值；當受試者給予否定回答時，有 $\frac{1}{2(1-\pi)}$ 的概率增加一個步長的值，有 $1 - \frac{1}{2(1-\pi)}$ 的概率降低一個步長的值。



貳、帶權重上下法

Kaernbach (1991) 同樣是在上下階梯法的基礎上提出帶權重上下法。由上下階梯法的迭代式 (見表達式(1)) 可得出, 當受試者給予肯定回答時, 下一次試驗的刺激強度一定降低一個步長的值; 當受試者給予否定回答時, 下一次試驗的刺激強度一定增加一個步長的值。因此, 在上下階梯法中, 每次增加或降低的刺激值是相等的, 即一個步長。Kaernbach (1991) 從此角度出發, 他認為儘管實驗中適測方法的步長固定不變, 但可以調整單次增加與降低的刺激值的權重, 使每次增加與降低的刺激值不相等。在此情況下, 為確保修改後的上下階梯法能收斂於心理計量函數的任一閾值, 他提出這種權重調整需要滿足以下表達式:

$$S_{up}\pi = S_{down}(1 - \pi) \quad (4)$$

其中, S_{up} 與 S_{down} 分別是單次增加的刺激強度值與降低的刺激強度值。在滿足此式的前提下, 對於任意介於 0 與 1 之間的目標概率, 下一次試驗的刺激強度可表達為:

$$X_{n+1} = X_n - \delta \left(\frac{Y_n - \pi}{\pi} \right) \quad (5)$$

此外, 對比此迭代式與德爾曼上下法的迭代式可發現, 帶權重上下法並沒有引入伯努利隨機變量。Kaernbach (1991) 解釋其原因是伯努利隨機變量帶來的隨機性可能會降低適測方法的效率而非像適測方法的目的所要求的提升其效率。



參、偏差硬幣投擲法

Durham與Flournoy (1995) 提出一種用於估計藥物治療中給予病人合適藥量的適測方法，它適用於各種嚴格遞增的函數且對函數形態無限制。該適測方法在上下階梯法的基礎上引入一枚有偏差的硬幣以增加隨機性，因此，這枚硬幣可表示為伯努利隨機變量；當選定的目標概率不低於0.5時，該方法計算的下一次試驗的刺激強度如下：

$$X_{n+1} = X_n - \delta(Y_n(S_\pi + 1) - 1) \quad (6)$$

其中， S_π 有 $\frac{1-\pi}{\pi}$ 的概率取值為1，有 $1 - \frac{1-\pi}{\pi}$ 的概率取值為0。因此，當受試者給予肯定回答時，下一次試驗的刺激強度有 $\frac{1-\pi}{\pi}$ 的概率會降低一個步長的值；有 $1 - \frac{1-\pi}{\pi}$ 的概率維持同樣的刺激強度。當受試者給予否定回答時，下一次試驗的刺激強度一定增加一個步長的值。而當選定的目標概率低於0.5時，下一次試驗的刺激強度如下：

$$X_{n+1} = X_n - \delta(1 - (1 - Y_n)(S_{1-\pi} + 1)) \quad (7)$$

其中， S_π 有 $\frac{\pi}{1-\pi}$ 的概率取值為1，有 $1 - \frac{\pi}{1-\pi}$ 的概率取值為0。由此得出：當受試者給予肯定回答時，下一次試驗的刺激強度一定降低一個步長的值；當受試者給予否定回答時，有 $1 - \frac{\pi}{1-\pi}$ 的概率維持同樣的刺激強度，下一次試驗的刺激強度有 $\frac{\pi}{1-\pi}$ 的概率增加一個步長的值。

在詳述其他問題前，本文在此先指出以上三種適測方法的一些異同。首先，偏差硬幣投擲法與帶權重上下法、德爾曼上下法的不同在於偏差硬幣投擲法並不是每次試驗的刺激強度都發生改變。其次，帶權重上下法與另兩者的不同在於它每次增加或下降的刺激值都是固定的，並沒有引入伯努利隨機變量，也因此它無需根據目標概率的位置轉換迭代式。而以上三種適測方法均能估計心理計量函數

上任意目標概率點的閾值且要求心理計量函數滿足其單調性假設。



肆、初始值、試驗數與終止條件

提出前述三種適測方法的學者均已證明，若經過無限次迭代，上述適測方法的迭代式（等式(2), (3), (5), (6), (7)）所得的刺激強度會最終收斂於閾值。若只經過有限次迭代，由其中任一迭代式可知，該迭代式最終收斂的刺激強度與兩個因素相關：第一次迭代時等式右邊的刺激強度（即初始值）、適測方法的步長。若步長過大，經過一定的迭代次數後，迭代式會在某一範圍內波動。但由於步長過大且恆定不變，在有限次內它無法收斂到閾值，只會來回擺動。反之，若步長過小，迭代式收斂速度過慢。類似地，若設置的初始值距離閾值很遠，則需要經過更多次地迭代才能最終收斂到閾值。綜上，初始值與步長的設定會影響同等迭代次數下的收斂結果，合理的步長與初始值能有效減少所需的迭代次數，因此本文將初始值與步長均列為實驗操作變量以探討步長與初始值對閾值估計結果影響。

此外，由於適測方法的迭代式可無限迭代，若沒有設定任何終止條件，實驗會無限進行下去。從迭代式來看，迭代的停止條件應是相鄰兩次試驗的刺激強度不再變化，但該條件需無限次迭代。這種情境下，實驗的終止條件可用刺激強度的差值來判斷，即迭代的最終估計值與閾值的偏差小於某個預定誤差值即可。該條件要求研究者已知閾值的位置，而閾值估計實驗目的是找到目標概率所對應的閾值，因此現實情況無法滿足此前提條件。另一種終止迭代式的條件是程序性終止：研究者預先設定實驗的試驗數上限，當實驗中的試驗次數達到設定值時自動終止。出此現實考量，本文用試驗數作為模擬實驗的終止條件：當模擬試驗次數達到預先設定試驗數值，單次模擬實驗終止。

另一方面，本文旨在探討在閾值估計實驗中引入回答信心量尺對適測方法的

影響；而適測方法的出發點是提高實驗效率且減少實驗資源的佔用。以此出發點，若某一適測方法在經過一定的試驗數後，其表現沒有明顯提升，在實驗中應採用最小試驗數以優化資源應用。此處適測方法的表現指的是經過某一階段後，適測方法迭代所得的刺激強度在某一合理的小範圍內波動，此後再迭代多次依舊不會躍出此範圍。而試驗數的多寡可能對不同適測方法影響程度不一，因此本文亦關注以不同試驗數作為終止條件時，各個適測方法在同等情境下的表現。

第三節 回答信心

雖然閾值估計實驗多採用二回答選項模式，此回答模式無法考量所有的研究情境所帶來的問題引起了學界的廣泛重視。此類回答模式引起的問題多出現於當試驗中比較刺激與標準刺激的差值很小的情境：受試者無法區分刺激間存在的差異，但「是」/「否」作業或「同」/「異」作業均要求受試者在每次試驗給出「是」或「否」的回答。因此，受於實驗本身的限制，受試者會被迫猜測答案回答。此外，「緒論」中已述，此類情境下猜測的答案會與受試者在十分確定的情境下給予的回答被同等對待，但實驗設計本身應避免受試者因實驗設計的限制而給予不可靠的答案。

針對此困境，Kaernbach (2001) 提出增加新的回答選項「不知道」形成三回答選項模式：當受試者主觀無法區分刺激的差異性時，可選該回答選項以避免猜測作答。Kaernbach (2001) 以二回答選項的帶權重上下法為基礎，推導出三回答選項的帶權重上下法並證明在閾值估計實驗裡，該適測方法的表現比上下階梯法的表現好。類似地，García-Pérez與Alcalá-Quintana (2017) 兩位學者也證明該方法在其他作業背景下的優越性。

儘管學者們已證明三回答選項模式比二回答選項模式在閾值估計實驗中表現

更佳，但無法打消其他學者對三回答選項模式另一方面的擔憂。Klein (2001) 分析在三回答選項的實驗情境裡，受試者應在信心程度很低時選擇「不知道」，但該選項實則同時涵蓋來自肯定回答與否定回答兩方面的不確定性。這種情況可能會讓受試者有更高的概率選擇「不知道」，其被選擇的比重可能遠高於其他兩者（「是」或「否」）。若出現此種情況，該實驗設計本身具有系統性偏誤，亦引發受試者出現回答偏誤，應進行修改。從現實角度考慮，由於難以確定受試者是否在實驗中出現對回答選項的傾向性選擇，故很難排除此類情況。

針對此憂慮，Klein (2001) 在三回答選項模式的基礎上，將新增加的「不知道」回答選項分割成爲兩個對立的低信心回答選項，即「低信心的是」與「低信心的否」；原來的「是」與「否」則對應修改爲「高信心的是」與「高信心的否」。由於將原先「不知道」的回答選項拆分成爲對立的低信心回答選項，四回答選項在一定程度上「均分」四個回答選項的概率，因此可以規避三回答選項可能帶來的回答偏誤。

儘管四回答選項的劃分方式十分合理，過度有限的回答選項數使其也具有一定的侷限性。以實驗設計角度而言，引入「低信心的是」與「低信心的否」回答選項的本意是給予受試者在其信心程度很低時更加自由的選擇權，但該處理方式無可避免地引入新的問題。若受試者處於「高信心」與「低信心」之間的過渡階段，四回答選項致使他只能被迫選擇其一，而這種情況本質與二回答選項所面臨的困境一致，僅僅是減輕了二回答選項的問題嚴重程度。

考慮到此類問題的影響，最直接的解決方式是將原來的四回答選項模式擴展爲多個回答選項，每一個回答選項代表一個特定的信心程度。具體而言，若研究者給予受試者 r 個回答選項，從 r_1 到 r_n 分別代表受試者的回答信心程度是「極有信心」、「較有信心」、「一般有信心」等依次降序排列。此時受試者所擁有的選擇自由度增加，所面臨的被迫選擇的困境也相應減少。

儘管多個信心回答選項的模式很合理，但有限個的信心回答選項仍存在一些

問題，而此問題與回答選項的個數多寡無關。為方便說明，此處先以最初的四選項回答舉例（即「高信心的是」與「高信心的否」、「低信心的是」與「低信心的否」），考慮以下兩種情境：其一，受試者本身對於「高信心」與「低信心」所對應的信心程度的標準極高：受試者僅在信心程度極高的情形下才認定自己是處於「高信心」狀態；或反之，受試者對自己的信心程度標準定義寬鬆，稍有信心的情形受試者就認為自己處於「高信心」的狀態。在第一種情境下，受試者給予「低信心的是」的回答與研究者提出的「低信心的是」的定義有一定落差，但研究者無法判定受試者給出的「低信心的是」是基於何種標準；反之亦然。

若觀察多信心回答選項模式可知，有限個的回答選項並無法解決此類問題。具體而言，假設在受試者心裡存在一個連續的回答信心量尺，不同受試者對於「極有信心」與「較有信心」之間的劃分值在該量尺上並不一致。這種個體差異性的存在，使每個受試者自我定義的「極有信心」與「較有信心」具有偏差。不僅如此，與四回答選項一樣，當兩名受試者在同等實驗條件下均給予「極有信心」的答案時，研究者無法獲知他們是否按照相同的信心標準作出該判定。這種潛在的差異性致使實驗中的回答信心選項的標準對於不同受試者而言並不一致。

若從另一角度思考此問題，受試者給予相同回答卻具有不同的信心標準，其根源在於每位受試者將其內心連續的回答信心量尺按自我標準切割成對應的多個信心回答選項。因此，若直接提供受試者一個連續的回答信心量尺，由受試者在該量尺上標定自己的信心程度。此時可以避免受試者信心切割標準的差異性，又可幫助研究者獲得更為詳實的數據。

儘管連續的回答信心量尺解決了上述問題，但這並不是多信心回答選項模式碰到的唯一問題。在「適測方法」部分已述，現階段的適測方法主要適用於二選項回答模式，若引入多個信心回答選項，則需按照Kaernbach（2001）引入三回答選項時修改適測方法的迭代式類似的做法，相應地修改應用於多個信心回答選項的適測方法。若研究者此時對回答選項進行增刪，應用於原先多個回答選項的

適測方法也要作出更動。此外，對於不同的適測方法，其所需的更動也不盡相同。

綜上兩個問題，理想狀況是保留連續的回答信心量尺以避免受試者對信心標準的差異定義，同時盡量維持原有適測方法不作更動。基於此目的，Hsu與Chin（2014）提出新的解決方式：他們建議給予受試者一個連續的回答信心量尺讓受試者在此量尺上作答。在每次試驗裡，受試者由原先回答「是」或「否」，轉為在回答信心量尺上回答他對自己主觀判斷的信心程度。

另一方面，研究者事先在量尺上選定一個常量值（本文定義為「截點」），此截點用於判定受試者的回答應按何種標準標定成二回答選項模式：若受試者回答的信心程度超過該截點，研究者認定該回答為肯定回答，相當於二選項回答的「是」；反之，研究者認定該回答為否定回答，相當於二選項回答的「否」。這種方式既保留了二回答選項模式，又解決了受試者對於不同的信心類別選項定義標準差異的問題。儘管這種做法無需更改適測方法本身，但它對適測方法在閾值估計實驗裡的應用影響還不明確。因此，本文從此出發，在前述三種適測方法中引入連續的回答信心量尺與截點，探討該回答機制對適測方法的影響。



第三章 模擬實驗

第一節 回答信心的構建

受試者在連續回答信心量尺上作答時，對不同刺激強度有不同的區辨能力。本文假設這種區辨能力可由回答信心的期望值體現：在兩個刺激的強度差值（刺激強度差=比較刺激-標準刺激）從負向到正向轉變時，受試者回答信心的期望值也隨之增加。另一方面，在「是」/「否」作業裡，當兩個刺激的強度差較大時，受試者能明顯判斷哪個刺激的強度大；若多次試驗保持該差值不變，此時因為受試者的信心程度很高，其信心程度的變化範圍很小。反之，若此時兩個刺激的強度差接近零，若保持該刺激差值不變且重復試驗多次，受試者難以判斷兩個刺激的大小差異。在此情形下，受試者對自己給出的回答極度不確定，他給予的信心程度變化範圍較大。因此，對於不同的刺激強度（以下刺激強度均指代刺激強度差值，即刺激強度差=比較刺激-標準刺激），受試者信心程度的變化範圍並不相同，而且回答信心變化範圍的大小亦反映受試者對自己答案的不確定程度。

既然不同刺激強度下回答信心的期望值不同且變化幅度也不一致，若將受試者在整個實驗中的回答信心按照刺激強度劃分，每個刺激強度所對應的回答信心應形成一個分布。因此，本文假定每一個固定的刺激強度有它對應的回答信心分布，該分布僅由刺激強度決定。回答信心分布與刺激強度的對應關係固定且唯一：每次試驗時，受試者從這個試驗的刺激強度所對應的回答信心分布中抽取他的回答。當下次試驗的刺激強度發生變化時，受試者切換到對應的回答信心分布，同樣從中抽取他的回答。

從不同刺激強度間的差異看，儘管不同刺激強度對應不同的回答信心分布，但同一個實驗裡這些回答信心分布應具有相同的函數形態。另外，給予受試者的連續回答信心量尺應有界限，本文在此設定該回答信心量尺的下界是零，上界是一，數值從小到大代表信心程度的增加。因此，本文模擬建模時假設回答信心分布的函數形態是截斷常態分布（truncated normal distribution），其下界是零，上界是一。因為不同的刺激強度所對應回答信心的期望值與變化範圍不同，在此期望值即截斷常態分布的平均數，變化範圍由截斷常態分布的標準差表示。

綜上而言，在「是」/「否」作業裡，當刺激強度上升時，回答信心的期望值會隨之增加，相應地代表該期望值的截斷常態分布函數的平均數亦隨之增加；因此截斷常態分布函數的平均數與刺激強度應成正相關，它們的函數關係設定為累積邏輯函數（logistic function）。又因在如下的實驗設計方式中，受試者不知比較刺激與標準刺激的位置，他只能主觀感受區辨兩個刺激的難度會隨兩者的差值的減小而增大，相應地回答信心的變化範圍亦隨區辨難度的增加而增加。因此，在同樣的刺激變化過程中，相應地截斷常態分布函數的標準差會先增大而後減小。根據這些假設，代表回答信心分布的截斷常態分布函數的平均值與標準差由如下兩個等式決定：

$$\text{mean of response confidence} = \text{cumulative logistic function}(i, 0, 0.35) \quad (8)$$

$$\text{standard deviation of response confidence} = \sqrt{0.25u(1-u)} \quad (9)$$

其中， i 是刺激強度（刺激強度=比較刺激-標準刺激）， u 是截斷常態分布函數的平均值（即等式(8)左邊）。

由上述兩個等式可知，刺激強度為零時，截斷常態分布函數的平均值是0.5，回答信心分布的標準差最大。此時兩個刺激的差值為零，受試者無法分辨左右刺激的大小差異，無論他給出的信心程度是多少，此時他對該回答的不確

定程度最大，代表這種不確定性的標準差也最大。而本文已假設回答信心的期望值應與刺激強度呈正相關且刺激強度由負無窮到正無窮變化：當刺激強度為零時（比較刺激與標準刺激強度一樣），信心程度的期望值相應地位於回答信心量尺的中點。

另一方面，若結合等式(8)與等式(9)可知，當比較刺激與標準刺激的刺激差值（刺激差值=比較刺激-標準刺激）具有相同的絕對值時，它們對應的回答信心分布函數（即截斷常態分布函數）具有相同的標準差；而且它們的平均數關於0.5對稱，即關於回答信心量尺的中點對稱。因此，在這種情境下，這兩個回答信心分布本身也關於0.5對稱。在此需說明，這種對稱性的產生是因本模擬實驗沒有考慮受試者的回答偏誤（即受試者本身不具有左右偏向性）；且根據實驗設計比較刺激與標準刺激的位置隨機，因此，具有相同絕對值的正負刺激強度，對受試者而言區辨難度是一樣的（因為它們的差值一樣）。因而，這種情況下它們的回答信心分布應具有相同的標準差。

上述回答信心分布設定（即等式(8)與等式(9)）是模擬信心實驗的基礎，以下詳述其過程。前文曾提及研究者會在回答信心量尺上預先選定一個固定的截點，在模擬實驗的過程中，該截點固定不變。因為每個刺激強度有其對應的回答信心分布，該截點在不同回答信心分布上的累積函數值也不同。因此，在固定截點的條件下，每個刺激強度亦有其對應的累積函數值（以下稱為累積信心函數值）。

在模擬實驗裡，這個累積信心函數值用於評定模擬的受試者回答應被認定為「是」或「否」。每次試驗從區間為 $[0, 1]$ 的均勻分布中隨機抽取數值作為受試者的模擬回答；若該模擬回答超過該刺激強度對應的累積信心函數值，則認定為二選項回答中的「是」，反之認定為二選項回答中的「否」。然後將該回答結果帶入到上述三種適測方法的迭代式裡計算下一次試驗的刺激強度。該過程裡回答結

果的設定可表述為如下等式：

$$Y_n = 1 \quad \text{when } \text{simulated response} \geq \text{cumulative confidence} \\ = 0 \quad \text{otherwise} \quad (10)$$



從模擬實驗做法可知， $Y_n = 1$ 等同於受試者在傳統的二選項回答實驗中回答「是」，且受試者在傳統實驗中回答「是」的概率與刺激強度的關係構成心理計量函數。當引入連續的回答信心量尺後，受試者在實驗中回答「是」的概率可通過回答信心分布的累積函數計算。因此，在固定截點的條件下，引入回答信心量尺的心理計量函數可表示成：

probability of psychometric function =

$$1 - \text{cumulative confidence}(\text{cutoff}, u, \sqrt{0.25u(1-u)}) \quad (11)$$

其中，*cutoff*是截點，*u*是截斷常態分布函數的平均值（即等式(8)左邊），並且 $\sqrt{0.25u(1-u)}$ 是截斷常態分布函數的標準差（即等式(9)左邊）。由該等式(11)可知，若截點固定不變，它對應一條心理計量函數；若截點發生變化，因為對應的累計信心函數值相應地發生變化，它會得到對應的心理計量函數。因此，每個位於回答信心量尺範圍內的截點有其對應的心理計量函數，這種對應關係唯一且均可通過上述等式計算。

第二節 3D/2D繪圖

本文已闡述刺激強度、回答信心與心理計量函數之間的關係，以下將以3D模型（見圖1）及其2D圖示（見圖2、圖3、圖4）拆分加以說明。3D模型可拆分為回答信心分布、刺激強度與回答信心期望值的關係、以及心理計量函數

三部分，其3D圖示（見圖 1）中每部分的對應關係可概述為：首先，已知每個刺激強度會有其對應的回答信心分布，三個截斷常態分布函數分別是刺激強度為-2,0,2的回答信心分布，且該三個刺激強度分別代表比較刺激小於標準刺激、比較刺激等於標準刺激、比較刺激大於標準刺激的情境（3D模型拆分後的2D圖見圖 2）。其次，最底部的綠線（標記為mean_confidence）描繪的是刺激強度與回答信心分布的期望值的關係，即等式(8)的關係；但此處參數均有所放大以便繪圖的簡明性（3D模型拆分後的2D圖見圖 3）。

此外，由等式(11)可知，每一個固定的截點對應一條心理計量函數。又因為研究者可以取連續的回答信心量尺上的任意一點作為實驗的截點，所有可選擇截點所對應的心理計量函數會形成一個連續曲面，即3D模型中的黃-紫色曲面。若在3D模型中的回答信心的軸線上選取一個截點，在該截點橫切一個平面並且該平面平行於刺激強度的軸；該平面與曲面的交線即該截點對應的心理計量函數，即等式(11)的關係（3D模型拆分後的2D圖見圖 4）。

回答信心分布函數所對應的2D圖切面圖如圖 2所示：紅線、綠線、藍線分別代表刺激強度為-2, 0, 2的截斷分布函數。從圖 2可知，當刺激強度增大時（紅線-綠線-藍線），截斷分布函數的平均值隨之上升，對應地回答信心的期望值也隨之增加；這與等式(8)相對應。另一方面，刺激強度增大時，其對應的截斷分布函數的散度先擴大後減小（如下圖紅線、綠線、藍線分別代表的截斷分布函數的標準差先增大後減小），符合等式(9)截斷分布函數標準差的變化，它們所對應的是受試者在這種變化過程中對於回答的不確定程度的變化。

此外，從等式(8)與等式(9)的性質可知，當正負兩組刺激強度具有相同的絕對值時（即比較刺激a-標準刺激b=標準刺激b-比較刺激c），它們自身關於0.5（即回答信心量尺的中點）對稱。從圖 2可知，紅線的截斷常態分布函數與藍線的截斷常態分布函數，它們所對應的刺激具有相同的絕對值，因此這兩個回答信心分布關於0.5對稱。而當標準刺激與比較刺激相等時（圖中綠線），因此

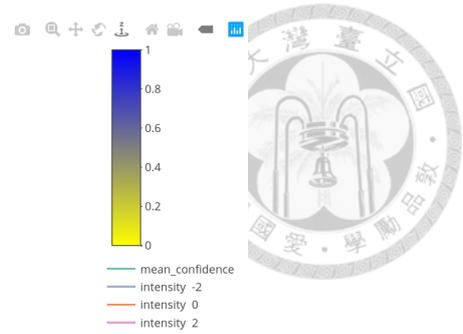
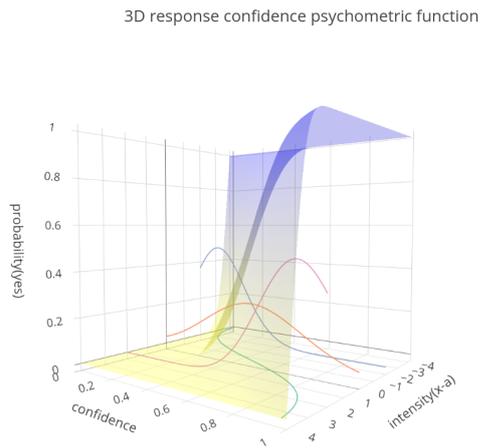
時受試者無法判斷兩個刺激的長度差異，他選擇左右刺激的概率均等，對應的回答信心分布函數亦會關於它自身對稱。

刺激強度與回答信心期望值的關係如圖 3 所示，圖中回答信心的期望值（即截斷分布函數的平均值）是刺激強度的累積邏輯函數（見等式(8)）。由圖 3 可知，回答信心的期望值與刺激強度成正相關，並且回答信心期望值的變化範圍與回答信心量尺的界限範圍一致。

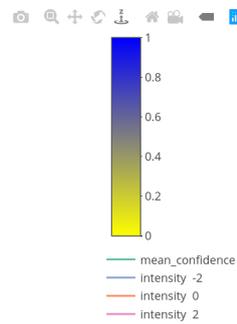
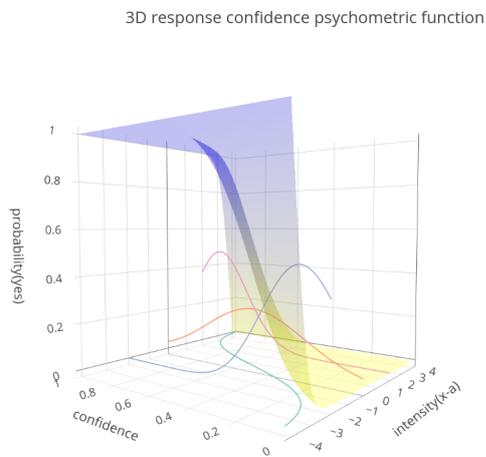
最後對 3D 模型（見圖 1）的曲面部分進行拆分（見圖 4），本文選取截點值是 0.2, 0.5, 0.8，在 3D 模型中對應地取其豎直切面。這三個切面與曲面的三條交線如圖 4 所示，它們分別是三個截點對應的心理計量函數。

由圖 4 可知，當截點是 0.2 時，受試者的回答信心程度不小於 0.2 即認定為「是」；若截點是 0.8 時，此時受試者的回答信心程度需達到 0.8 才會被認定為「是」。對比截點是 0.2 與截點是 0.8 這兩種實驗條件，在比較刺激比標準刺激小的情境裡，若受試者給予某個回答信心且該回答信心超過 0.8，在同樣情境裡該回答必然也超過 0.2；反之，若回答信心超過 0.2，在同樣情境裡卻不一定超過 0.8 的截點。因此，在同樣刺激強度下的高截點實驗裡，受試者回答信心被認定為「是」的概率會比低截點的實驗裡受試者的回答信心被認定為「是」的概率低。因此在相同的刺激強度下，低截點的心理計量函數值會高於高截點對應的心理計量函數值。從目標概率而言，相同目標概率的截點越小對應其閾值越小。

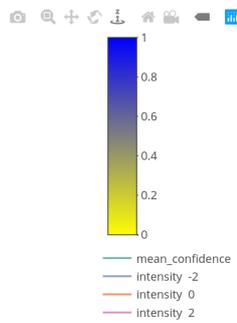
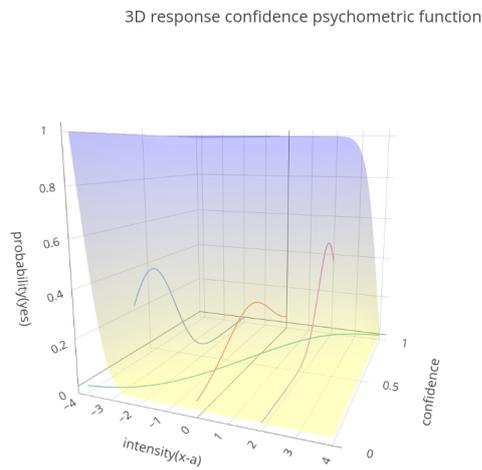
另一方面，由圖 4 可知，儘管不同截點有其對應的心理計量函數，但這些心理計量函數具有相同漸進線。而且，這些心理計量函數間的距離並不一定「等距」：比如截點 0.2 與截點 0.8 到截點 0.5 的信心距離相同，但它們的心理計量函數並不一定具有相同的函數斜度。進一步地說，若回答信心分布的設定稍作變化，在相同目標概率的條件下，截點 0.2 與截點 0.8 所對應的心理計量函數閾值不一定關於中間截點所對應的閾值對稱。



(a) 截圖 1



(b) 截圖 2



(c) 截圖 3

圖 1：3D模型截圖

註：圖中底部的綠線 (mean_confidence) 是以等式(8)為基礎，將其中的標準差放大繪得。
此3D繪圖模型由R庫Plotly繪製。

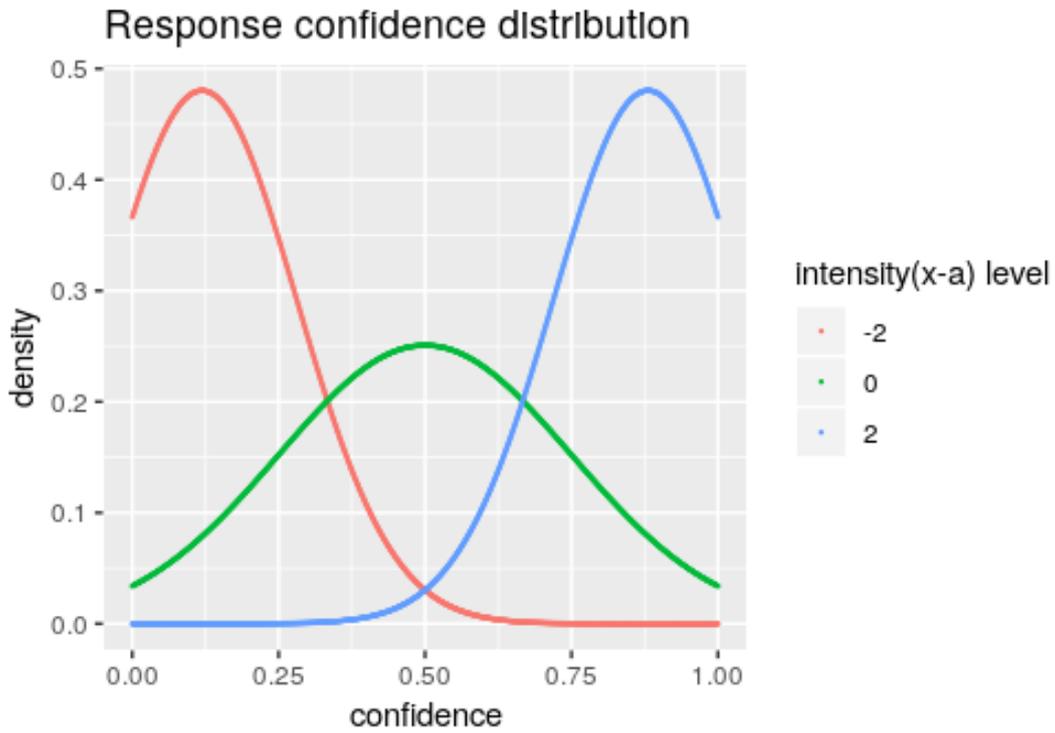


圖 2：2D分圖-回答信心分布函數的概率密度函數

註：右側圖例表示比較刺激 (x) 與標準刺激 (a) 的強度差值。圖中的紅線、綠線、藍線分別是刺激強度為-2, 0, 2對應的回答信心分布函數的概率密度函數。橫軸代表回答信心量尺，它的範圍是[0,1]；縱軸代表回答信心函數的函數密度。

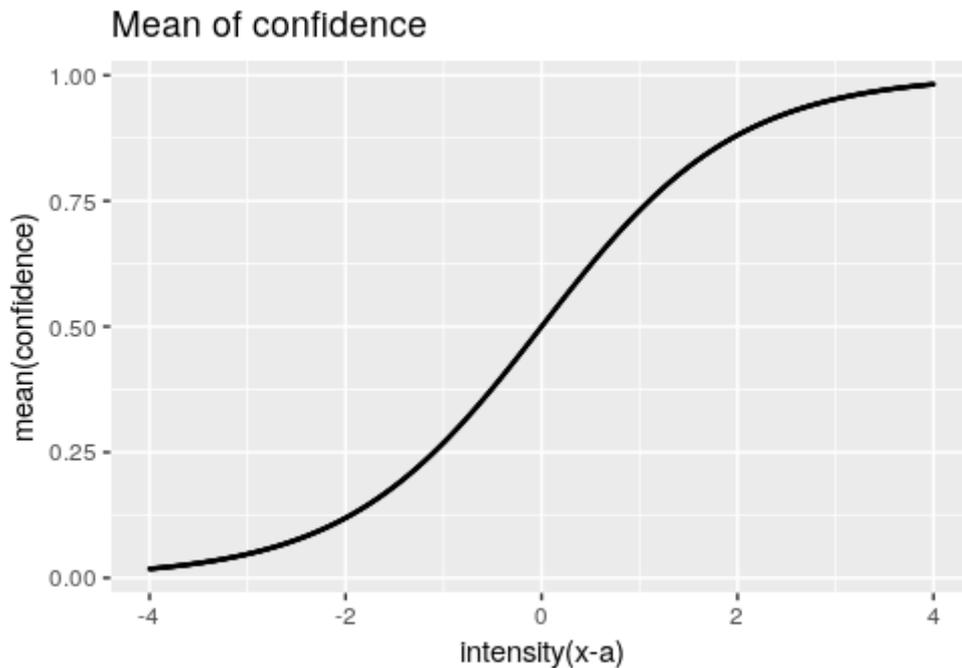


圖 3：2D分圖-回答信心的期望值與刺激強度的關係

註：為繪圖結果的清楚性，此處該累計邏輯函數的參數和等式(1)的參數並不一致。

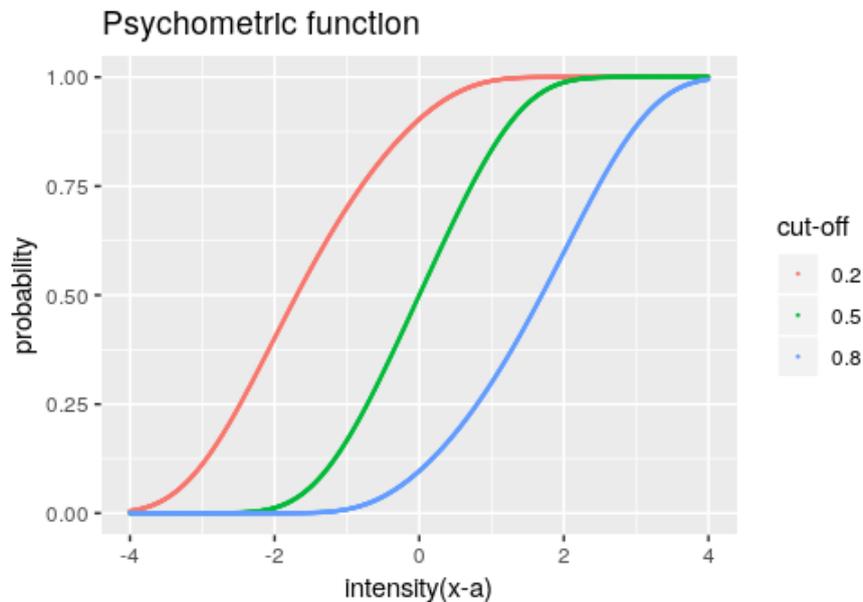


圖 4：2D分圖-心理計量函數與其對應的截點關係

註：截點為0.2, 0.5, 0.8時對應的心理計量函數。

第三節 適測方法、截點、目標概率、步長、初始值、試驗數、模 擬次數

本模擬實驗涉及到以下實驗變量：適測方法、截點、目標概率、步長、初始值、試驗數。因為該模擬實驗目的是估計給定目標概率對應閾值，因此預先選定兩個目標概率是0.75和0.5。其中，0.75是心理物理學研究中常用的目標概率，而本模擬實驗以0.5的目標概率為基準點對照引入回答信心量尺對適測方法的根本影響。模擬實驗採用的適測方法是德爾曼上下法、偏差硬幣投擲法和帶權重的上下法。另外，在連續的回答信心量尺上選取截點可分為三種情境：低截點、中間截點、高截點，在此相應地選取三個截點：0.2, 0.5, 0.8分別代表低截點、中間截點、高截點的情境。

本文已述模擬實驗的終止條件由試驗數決定，且實驗在進行至一定的試驗數後，適測方法的表現可能不會有明顯改善。因此，本模擬實驗在此選取的試驗數

範圍為50到150，以間隔25個試驗數為單位，分別模擬各個試驗數的閾值估計結果。其中，試驗數的範圍根據現實情況與模擬結果所選定，試驗數的上限是基於現實中對受試者體力的考量，若超過該試驗數，受試者的負擔較大。下限則是根據預先模擬實驗結果所決定，若小於該試驗數，所有適測方法表現均不如人意。

而在適測方法的迭代式中，初始值與步長都會影響到迭代效率。若步長過大，適測方法可能最後來回擺蕩無法收斂；若步長過小且初始值偏離閾值過遠，適測方法收斂速度過慢，需較多次的試驗才能收斂到閾值。因此本模擬實驗選取的步長範圍是0.04到0.20，間隔為0.04的所有步長。而初始值從它相對於閾值的位置可分為三類情境：低於閾值、在閾值附近、高於閾值¹。此外，不同的目標概率對應的閾值不同，而相同的目標概率在不同的截點下對應的閾值也不同。

因此，本模擬實驗按照上述三種情境選取對應的初始值，在目標概率為0.75的情境下，初始值設為-0.4, -0.28, 0.25, 0.87, 1。其中，中間三個初始值分別在截點為0.2, 0.5, 0.8對應的閾值附近，對其中任一截點而言，另外兩個截點的閾值相對它都屬於另外兩個情境。意即，在截點0.2的條件下，-0.28在它的閾值附近，而截點0.5對應的閾值0.25，及截點0.8對應的閾值0.87都是屬於高於閾值的情境；而-0.4則是低於閾值的情境。因此對每個截點而言，模擬實驗所選取的初始值範圍都涵蓋上述三種情境，進而可對比初始值的位置對閾值估計的影響。在目標概率為0.5的情境下，初始值設為-0.75, -0.59, 0, 0.59, 0.75；其設定方式與目標概率為0.75的設定方式相同。

綜上，本模擬實驗的每組實驗條件是由適測方法、截點、目標概率、步長、初始值、試驗數組合而成。在每組實驗條件下，同一個實驗會重復模擬一千次。因為模擬實驗是閾值估計實驗，以下考慮單個實驗（即在一組固定實驗條件下的單個實驗），它最終估計的刺激強度以如下方式計算：在整個實驗中，扣除最初

¹儘管在現實中研究者很少能知道閾值的位置，但在本模擬實驗中，閾值是可預先反向計算的。因為本模擬實驗可通過刺激強度、回答信心分布與心理計量函數間的關係求得心理計量函數的表達式（見等式(11)），只需確定目標概率，可採用牛頓法反向計算對應的刺激強度。

兩個反轉點，計算餘下反轉點的平均值作為最終的估計刺激強度，即估計閾值（estimated threshold）。

這種計算方法會遇到一種情況：在試驗數較小的條件下可能不具有足夠多的反轉點，即扣除最開始的兩個反轉點後，已無反轉點；或總共都不具有兩個反轉點。若出現此種情況，本模擬實驗採用最終的迭代值作為這次實驗最終的估計閾值，即最後一次迭代後得到的刺激強度。因此，在每組實驗條件下，重復一千次模擬後，該組實驗條件的最終估計閾值是這一千次實驗的估計閾值的中位數。此外，本模擬實驗將這組實驗條件的最終估計閾值帶入等式(11)計算其對應的估計目標概率（estimated target probability），並與該組實驗條件的目標概率進行對比以評估適測方法的表現。

第四節 均方根誤差與偏誤

雖然對比估計目標概率與真正目標概率的差值可評估不同實驗條件下適測方法的表現，但適測方法目的是估計閾值，因此在評估適測方法時仍需從閾值的角度評估。另一評估適測方法表現的途徑是衡量真正閾值（本文簡述「閾值」均指代真正閾值）與估計閾值的差距，即評估最後估計的刺激強度的準確度。不僅如此，適測方法亦以優化實驗資源提高效率為目標：Leek（2001）提出適測方法應該平衡效率最大化與精確度最大化間的關係。適測方法的效率指適測方法經過多少次迭代能收斂到閾值，即需要多少試驗數能收斂到閾值。精確度指最終的估計閾值與閾值的接近程度。而這兩個標準在衡量適測方法的表現時是非常重要的指標，因為它們是適測方法的最終目的。

統計學裡有兩種指標分別衡量這兩個標準：以均方根誤差評估效率，偏誤評估精確度。儘管標準誤與均方根誤差皆反映估計閾值的分散程度，但均方根誤差

反映在同組實驗條件下，一千個估計閾值關於真實閾值的分散程度；而標準誤反映它們關於估計閾值的平均值的分散程度。另一方面，偏誤是衡量估計閾值的期望值與閾值的偏差程度。在此採用以下等式計算均方根誤差（見等式(12)）和偏誤（見等式(13)），並以此評估適測方法的表現。

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (\hat{\theta} - \theta)^2}{N}} \quad (12)$$

$$bias = \frac{\sum_{i=1}^N \hat{\theta}}{N} - \theta \quad (13)$$

其中， N 是在同組實驗條件下重復模擬實驗的次數，即一千次。 $\hat{\theta}$ 是每一次實驗的估計閾值， θ 是在該實驗條件下的真正閾值。在此需要說明的是，不同的實驗條件組合，它們的均方根誤差和偏誤均獨立計算。本文在結果分析的部分呈現均方根誤差和偏誤的分析結果。



第四章 模擬結果

模擬實驗結果分為三部分討論：前述在模擬實驗最後會將估計閾值帶入心理計量函數（第三章等式(11)）求出對應的估計目標概率，在此呈現其估計目標概率並分析該結果。雖然本文亦呈現均方根誤差和偏誤的結果分析，借此評估各種適測方法的精確度與效率，但對比估計目標概率與目標概率的偏差可以最直觀地反映適測方法的精確度。而均方根誤差和偏誤則從刺激強度的層面對比各種實驗條件下不同適測方法的表現。這兩個層面都可看出各種因素（即步長、試驗數、初始值等）對不同適測方法在引入回答信心量尺後的影響差異。

第一節 估計目標概率圖

在估計目標圖（見圖 5）裡，每一個點代表將一千次估計閾值的中位數帶入等式(11)後得到的估計目標概率，每張估計目標概率圖上方羅列其對應的實驗條件依次為：適測方法、試驗數、截點以及目標概率。其中，右下方的圖例表示初始值，橫軸是步長，縱軸為每組實驗條件模擬一千次後的估計目標概率。

以下先陳述偏差硬幣投擲法的模擬結果。在步長較小的實驗條件下（步長為0.04或0.08，見圖 5），偏差硬幣投擲法出現估計偏誤，這種偏誤的方向由截點的位置所決定。不論目標概率的位置，低截點均會出現過高估計目標概率的結果；反之，高截點則會出現過低估計目標概率的結果。

在目標概率為0.75的模擬實驗裡（見圖 5(b)），高截點會導致多數實驗條件

均低估目標概率；且所有的估計點在增大步長時重合，表明增大步長與調整初始值均無法改變這種低估的趨勢。相反地，步長較小時若初始值的位置合適，它所估計的目標概率會更接近真實的目標概率。

另一方面，當步長增大時（見圖 5），目標概率的估計點開始重合，暗示初始值對目標概率收斂點的影響逐漸減小。這表明步長對目標概率估計準確度的影響比初始值位置的影響更大，因此偏差硬幣投擲法對步長的敏感度比對初始值的敏感度更高，它的估計準確度更容易受到步長的影响。

但在步長較小時，初始值嚴重影響估計目標概率的準確度：這種影響由初始值的位置與截點共同決定。在目標概率為0.75時（見圖 5(a)(b)），本模擬實驗設定低截點閾值附近的初始值為-0.28，高截點閾值附近的初始值為0.87。由圖 5(a)(b)可看出，低截點下小於閾值的初始值比高於閾值的初始值更有優勢；反之，高截點的條件下，高於閾值的初始值比低於閾值的初始值的估計結果更加準確。初始值的這種位置優勢在目標概率為0.5時（見圖 5(c)(d)）同樣保持。因此，在這種小步長的條件下，截點與初始值的位置會共同決定最後估計的目標概率的準確程度。

在步長較小時，若分別縱向對比（比較圖 5(a)(c)、比較圖 5(b)(d)）估計點的分散程度可發現，初始值的位置對目標概率為0.5的影響相較於對目標概率為0.75的影響較小，在目標概率為0.5時（見圖 5(c)(d)）估計結果亦更為準確。總體而言，在相同實驗條件下，偏差硬幣投擲法在目標概率為0.5時所估計的目標概率，比目標概率為0.75時所估計的目標概率更為準確。因此目標概率的位置亦影響偏差硬幣投擲法的表現。

步長較小時，偏差硬幣投擲法的這種估計偏誤亦體現在德爾曼上下法的結果裡。儘管德爾曼上下法在目標概率為0.75的條件下（見圖 6(a)(b)），當步長增大時整體都呈現過高估計的趨勢；但在步長較小的實驗條件下（步長為0.04或0.08），德爾曼上下法同樣受到截點位置的影響：在低截點的條件下高

估目標概率，在高截點的條件下低估目標概率。

不僅如此，該方法亦受到截點位置與初始值位置關係的共同作用（見圖 6(a)(b)）：在低截點下，初始值小於閾值比初始值大於閾值的條件估計更精確；反之，在高截點下，初始值大於閾值比初始值小於閾值的條件更有優勢。這種初始值的位置優勢亦體現在目標概率為0.5的情境裡（見圖 6(c)(d)）。

另一方面，圖 6亦表明步長與初始值的影響關係：具有不同初始值的估計點隨步長增大而開始靠近或重合，說明在德爾曼上下法裡，步長對其影響大於初始值的影響；這點與偏差硬幣投擲法的發現相同。而目標概率的位置同樣影響德爾曼上下法估計的準確性（比較圖 6(a)(c)，或比較圖 6(b)(d)）：在目標概率為0.5的條件下其估計準確度比在目標概率為0.75的條件下更好。

與德爾曼上下法在目標概率為0.75時整體高估目標概率的趨勢一樣（見圖 6(a)(b)），帶權重上下法亦表現出同樣的傾向（見圖 7(a)(b)）：隨步長的增大，整體結果均呈現稍微偏高估計目標概率；且估計點的逐漸重合也表明步長對其影響大於初始值對其結果準度的影響。類似地，在小步長的條件下，帶權重的上下法的估計準度亦受到截點的位置影響：在低截點時過高估計目標概率，在高截點時低估目標概率。

與此同時，與其他兩個適測方法一樣，截點位置與初始值的位置亦同步影響帶權重上下法在小步長條件下的估計準度（見圖 7）：在低截點的條件下，初始值小於閾值比初始值大於閾值的條件更具有優勢；反之，在高截點的條件下初始值大於閾值的模擬結果表現更好。這三種適測方法的相似亦體現在目標概率上，目標概率為0.5時（見圖 7(c)(d)），帶權重上下法表現比目標概率為0.75時（見圖 7(a)(b)）更好。

綜上，儘管德爾曼上下法在相同的實驗條件下估計目標概率的準確度不如帶權重上下法（見圖 6(a)(b)、圖 7(a)(b)），但它們在截點位置、步長、初始值等因素的影響下會呈現相同的偏誤趨勢。而偏差硬幣投擲法則與這兩者有些不同：

對比目標概率為0.75（見圖 5(a)(b)、圖 6(a)(b)、圖 7(a)(b)）估計結果可知：隨步長增大，偏差硬幣投擲法會出現向下低估的情況；但德爾曼上下法與帶權重上下法會出現向上高估目標概率的情況，其中德爾曼上下法尤為突出。

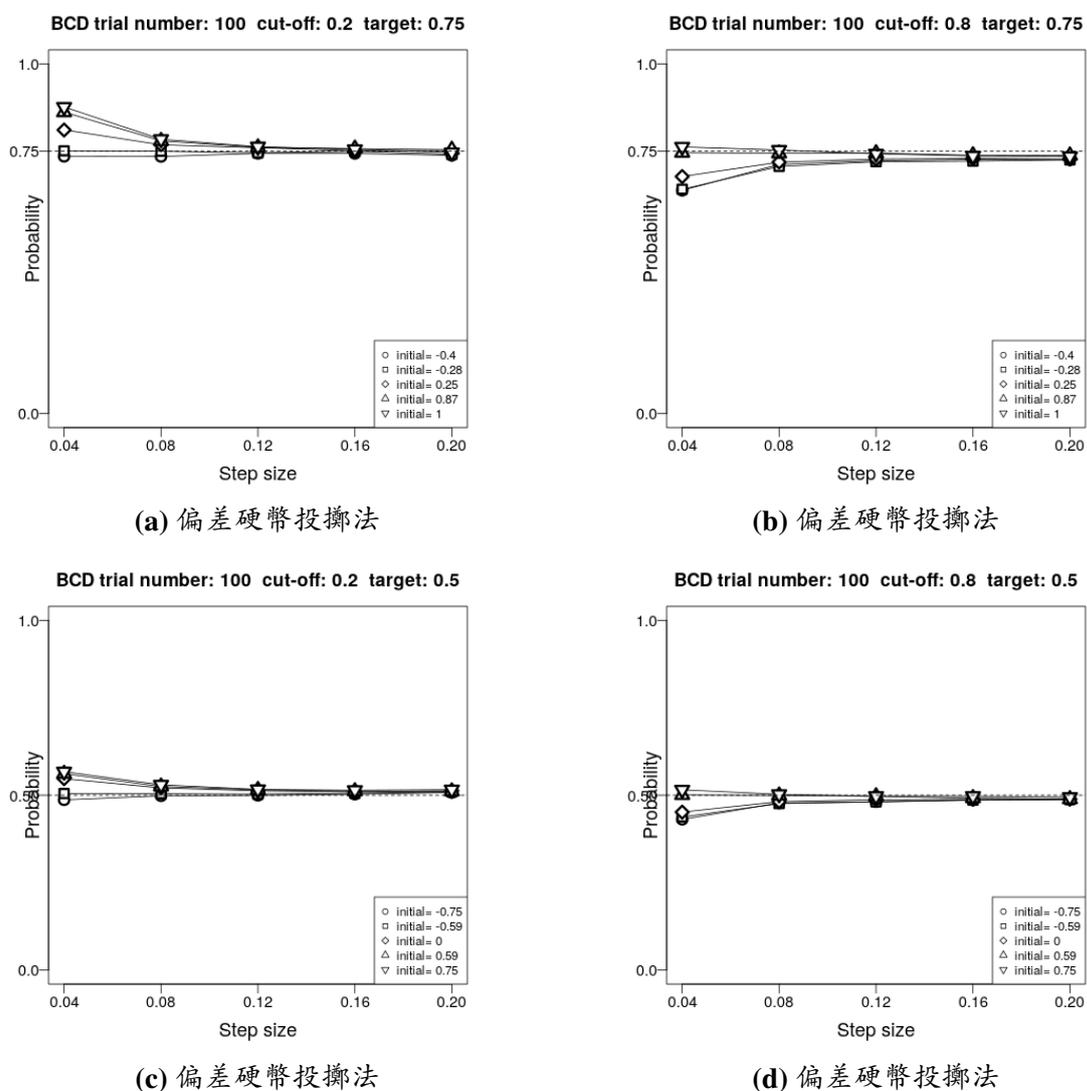


圖 5：偏差硬幣投擲法-估計目標概率圖

註：目標概率分別為0.75與0.5時，高截點與低截點對應的估計目標概率圖，其適測方法為偏差硬幣投擲法。

此外，對比三種適測方法在同樣實驗條件下可知（可參考圖 5(a)、圖 6(a)、圖 7(a)），德爾曼上下法的估計點的分散程度遠大於其它兩種適測方法，表明它受到初始值的影響比其它兩種適測方法所受到的影響嚴重。而步長太大的條件下（步長為0.16或0.20，可參考圖 5(a)、圖 6(a)、圖 7(a)），德爾曼上下法估計概率的偏離程度比其他兩種適測方法的偏離程度大。總體而言，偏差硬幣投擲法與

帶權重上下法的表現較好，德爾曼上下法在相同實驗條件下表現較弱於前兩者。

需要說明的是，本模擬實驗亦以試驗數為實驗操作變量，雖然僅呈現試驗數為100的實驗條件下適測方法的表現，但這些結果在試驗數較小的實驗條件下依舊存在且更為突出。在此選取試驗數為100主要基於以下考慮：由前述3D模型可知（見圖 1），相比於傳統的閾值估計實驗，本文新增回答信心這一維度，且在原刺激強度與心理計量函數的維度上新增一維度所需試驗數亦增加。因為回答信心分布隨刺激強度變化而不斷切換，實驗中適測方法在不同刺激強度間切換時回答信心分布可能變化很大，因此，儘管在傳統的閾值估計實驗裡採用100個試驗代價較高，但增加回答信心維度後本文採用100個試驗可能已是較少不負累受試者的試驗數。最後，因篇幅限制且避免重復贅述，本文將以均方根誤差和偏誤的結果呈現在不同試驗數的實驗條件下各個適測方法的表現。

第二節 均方根誤差

在此呈現不同實驗條件下一千次模擬的均方根誤差結果。在均方根誤差圖裡，每一個點均代表一千次模擬的均方根誤差（按第三章等式(12)計算）。圖例中顏色表示初始值，符號表示步長，均方根誤差圖以試驗數為橫軸，以均方根誤差為縱軸。每張圖左上標明其適測方法、目標概率及截點。

由圖 8(a)(c)(e)可知，當目標概率為0.75時，若其他實驗條件一致，最小的步長（0.04）具有最大與最小的均方根誤差（圖中實心原點）：當初始值位於閾值附近時具有最小均方根誤差；而初始值離閾值較遠時具有最大均方根誤差。此亦驗證估計概率圖在小步長條件下估計點分散的結果。但截點位置仍帶來不同影響：在中間截點（見圖 8(c)）的實驗條件裡，只有在閾值附近的初始值具有最小均方根誤差。在低截點（見圖 8(a)）時，其閾值附近的初始值與小於閾值

的初始值都具有較小的均方根誤差，且兩者相差無幾。類似地，在高截點（見圖 8(e)）時，初始值大於其閾值與位於閾值附近所得到的均方根誤差結果相近。這亦與前述截點位置與初始值共同決定估計概率準度的結果相一致。而在目標概率為0.5時（見圖 8(b)(d)(f)），截點位置與初始值亦有此關聯性。

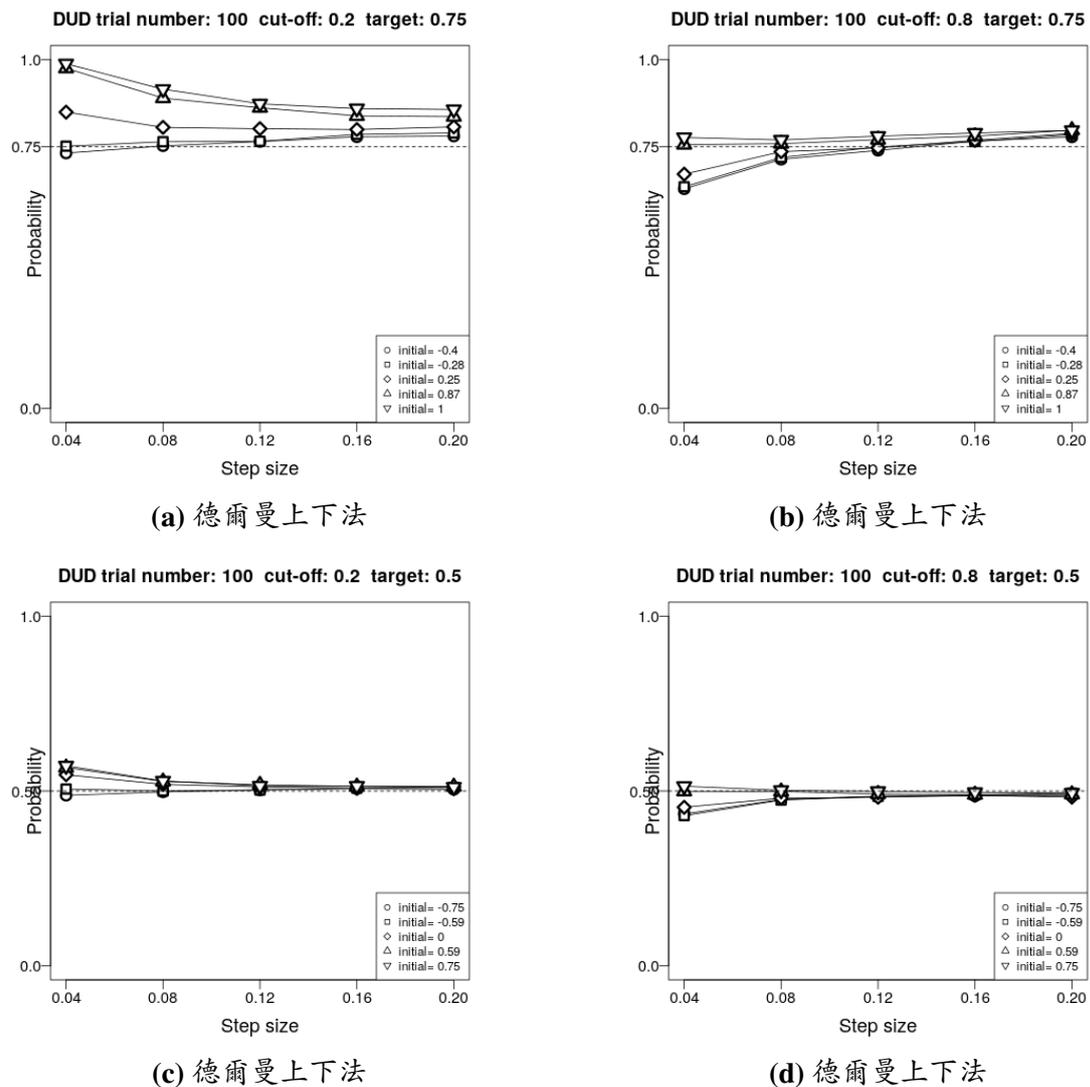


圖 6：德爾曼上下法-估計目標概率圖

註：目標概率分別為0.75與0.5時，高截點與低截點對應的估計目標概率圖，其適測方法為德爾曼上下法。

另一方面，對單一組實驗條件而言（圖中任意單一折線），若圖中折線平緩，表明增加試驗數無法有效減小均方根誤差；反之，若折線陡然下降，則增加試驗數使均方根誤差顯著降低。觀察均方根誤差的減小速率與試驗數的關係具有現實意義：研究者可根據研究所容忍的均方根誤差範圍，以及實驗資源的考量決

定合適的試驗數。若在某一組實驗條件下，試驗數增大時均方根誤差基本無太大變化且均方根誤差值低，則表明該組實驗條件在此適測方法的優越性；意即該組實驗條件具有理想的初始值、截點與步長。

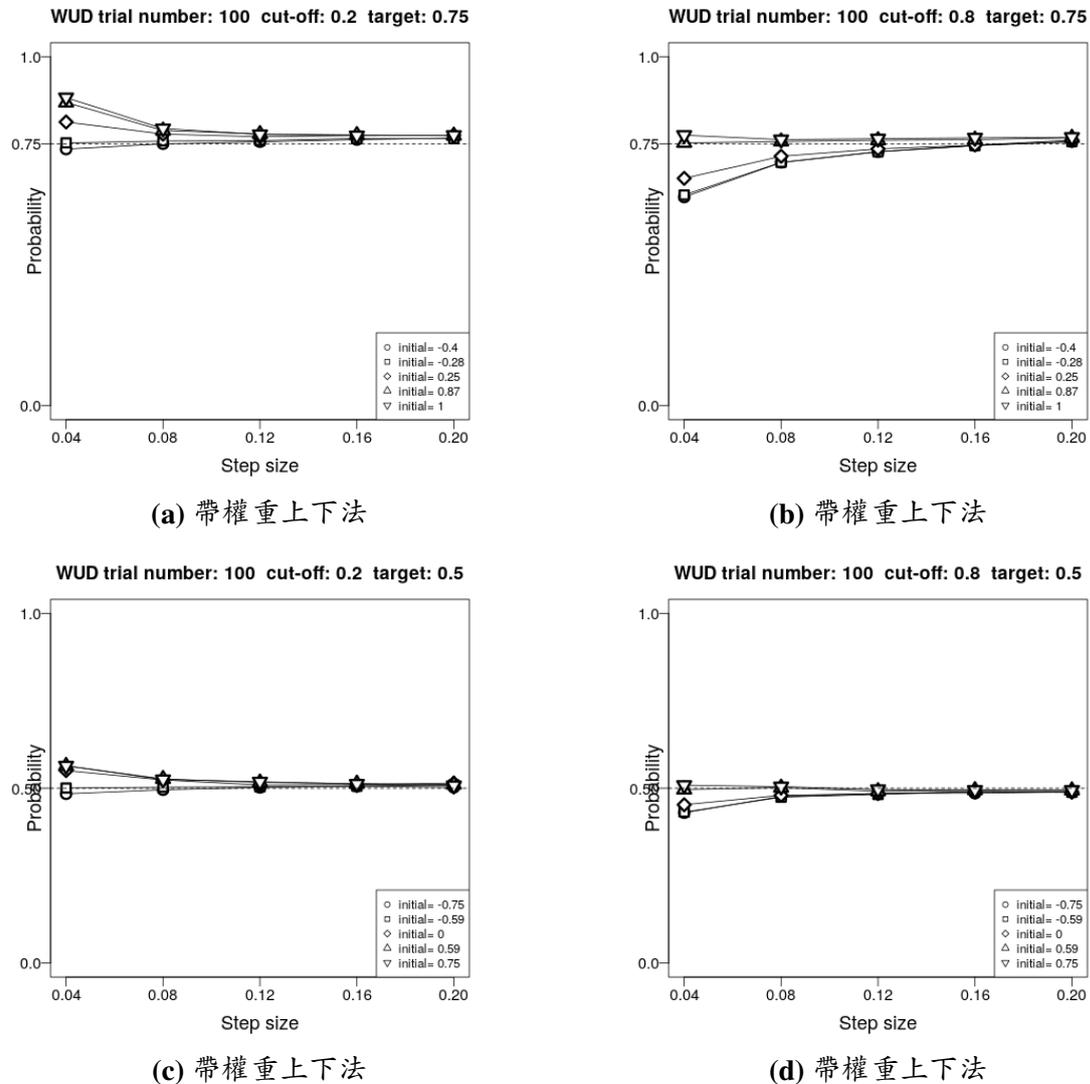


圖 7：帶權重上下法-估計目標概率圖

註：目標概率分別為0.75與0.5時，高截點與低截點對應的估計目標概率圖，其適測方法為帶權重上下法。

圖 8(a)(c)(e)中（目標概率為0.75）若除去均方根誤差偏離較大的幾個實驗條件（即步長最小且初始值遠偏離閾值的條件），其它實驗條件所對應的均方根誤差隨試驗數增大而逐步疊合，此表明當實驗超過一定的試驗數後，步長與初始值對其均方根誤差的影響逐步減小。對比中間截點、高截點（見圖 8(c)(e)）與低截點（見圖 8(a)）的均方根誤差折線可知，低截點的均方根誤差折線具有較大的下

降速率，表明其受試驗數的影響大於另外兩個截點。其中，隨試驗數增大，在中間截點與高截點時，聚簇折線的均方根誤差實則減小很少（此處圖 8(c)(e)具有不同的縱軸長度）。這表明在這兩種情境（高截點情境與中間截點情境）下，偏差硬幣投擲法的估計準度受到試驗數的影響遠小於步長與初始值。因此，研究者若根據實驗目的與實驗的可容許誤差範圍相應地減少試驗數，這種妥協對最終結果的影響比修改步長或初始值的影響更小。在低截點時，其均方根誤差受試驗數影響較大，此時需根據實驗情境適當選取試驗數。

最後，在目標概率為0.75時，對比三個截點的均方根誤差範圍（見圖 8(a)(c)(e)）可發現，低截點的最大均方根誤差遠大於其它兩者的最大均方根誤差，而中間截點與高截點的最大均方根誤差則較為接近。該結果表明，截點的相對位置與其均方根誤差的量級並不存在線性對應關係。意即，低截點與高截點到中間截點的距離相等，但高截點所對應的均方根誤差的上限遠小於低截點所對應的均方根誤差；這種關係本身是不對稱的。而目標概率為0.5時（見圖 8(b)(d)(f)）卻不具有這種不對稱的特點：高截點所對應的均方根誤差範圍與低截點所對應的均方根誤差範圍接近（見圖 8(b)(f)），但它們的誤差浮動範圍都大於中間截點所對應的均方根誤差範圍（見圖 8(d)）。

另一方面，儘管在此僅展示偏差硬幣投擲法的均方根誤差結果，但帶權重上下法與偏差硬幣投擲法具有類似的結果，在此不再贅述（附錄一列出三種適測方法在各組實驗條件下其均方根誤差的最大/最小值）。而德爾曼上下法與這兩種適測方法的結果稍有不同，在此僅詳述其差異性。具體而言，在目標概率為0.75時（見圖 9(a)(b)），德爾曼上下法在中間截點與高截點的均方根誤差與其他兩種適測方法的均方根誤差有稍微不同的趨勢。在偏差硬幣投擲法（見圖 8(c)(e)）折線重合的部分，均方根誤差隨試驗數增大而減小，因此折線整體趨勢由左上至右下斜向延伸；然而在德爾曼上下法（見圖 9(a)(b)）折線重合的部分隨試驗數增大時，它們具有明顯的起伏波動（均方根誤差有時反而會上升而非下

降)，即呈現波浪形態。

仔細觀察這些波浪狀折線所對應的實驗條件，它們均出現在步長較大(0.20)的條件裡。但根據德爾曼上下法的迭代式，試驗數增大應促使均方根誤差減小，因此這種波浪狀折線的存在表明適測方法已開始在閾值附近「擺蕩」(擺蕩指適測方法在閾值周圍來回搖擺，但無法擺到閾值點)。此結果表明德爾曼上下法對步長的敏感性較高，且傾向於較小步長；若步長過大則可能無法收斂。與之不同，偏差硬幣投擲法與帶權重上下法沒有明顯的波浪狀折線，表明它們對步長的敏感度較低，且在較大步長的條件下，初始值的影響幾近消失；因此，這兩種適測方法需採用較大步長。

最後，由附錄一可知，在同組實驗條件下，均方根誤差在目標概率為0.75下波動範圍相較於目標概率為0.5的波動範圍大，此表明三種適測方法在目標概率為0.5的條件下表現更好。這與估計目標概率的結果一致。

第三節 偏誤

以下簡述模擬實驗中各個適測方法的偏誤結果。在偏誤圖裡，每一個點均代表一千次模擬的偏誤(按第三章等式(13)計算)；其中，圖例的顏色表示初始值，符號表示步長，偏誤圖以試驗數為橫軸，以偏誤為縱軸。每張圖左上標明其適測方法、目標概率及截點。它與均方根誤差不同：均方根誤差僅有非負值，偏誤值可正可負。

因在誤差分析裡，德爾曼上下法、偏差硬幣投擲法與帶權重上下法結果近似，上述均方根誤差結果以偏差硬幣投擲法為主，在此以帶權重上下法為主呈現偏誤的分析結果。目標概率為0.75時(見圖 10(a)(c)(e))，偏誤的方向與截點位置相關：在低截點時，多數實驗條件裡帶權重上下法的估計閾值都高於真正閾

值，呈現高估閾值的結果；在中間截點時，高估與低估閾值的結果同時存在；在高截點時，則呈現低估閾值的結果。隨試驗數的增加，相同實驗條件所對應的偏誤均減小，因而呈現喇叭狀或半喇叭狀。這個結果與估計目標概率的結果一致（見圖 7(a)(b)）。類似地，目標概率為0.5時（見圖 10(b)(d)(f)）亦呈現上述趨勢。

而步長最小（0.04）時（見圖 10(a)(c)(e)或見圖 10(b)(d)(f)），初始值小於閾值導致偏誤為負，此時估計閾值小於真正閾值；若初始值大於閾值則估計閾值高於閾值；當初始值在閾值附近，偏誤近乎為零。這種趨勢與截點位置無關，但初始值與閾值的距離影響偏誤的絕對值：當初始值距離閾值越遠，其偏誤的絕對值越大，偏誤的正負性由初始值的位置決定（小於閾值或大於閾值）。

對比目標概率為0.5與目標概率為0.75兩種條件（分別比較圖 10(a)(b)、圖 10(c)(d)、圖 10(e)(f)）可知，在同等實驗條件下，目標概率為0.5所對應的偏誤小於目標概率為0.75所對應的偏誤。此表明帶權重上下法在目標概率為0.5時估計閾值比目標概率為0.75時準確，這也與前述估計目標概率的結果一致。

另一方面，在目標概率為0.75的條件下（見圖 10(a)(c)(e)），偏誤的波動範圍與截點的位置關係也不是「線性」的：低截點所對應偏誤的變化範圍（見圖 10(a)）遠大於中間截點、高截點所對應偏誤的變化範圍（見圖 10(c)(e)）。而目標概率為0.5的條件則不會出現這種不對稱性：低截點所對應偏誤的變化範圍（見圖 10(b)）與高截點所對應偏誤的變化範圍（見圖 10(f)）接近，且它們偏誤的變化範圍都大於中間截點的偏誤變化範圍。

此外，在圖 10(a)(c)(e)中若去除步長較小（0.04或0.08）的折線，其他實驗條件所對應的折線隨試驗數的增加偏誤變化較小；這與均方根誤差稍有不同：試驗數的增加仍在一定程度上減小均方根誤差值。因此，試驗數的增加能更有效地減小均方根誤差而非減小偏誤。從偏誤的結果而言，合適的步長與初始值組合能最小化偏誤。本文前述偏誤衡量適測方法估計閾值的精確度，均方根誤差衡量適測

方法的效率。因此，在同等試驗數的條件下，最小均方根誤差表明該實驗條件最有效率；偏誤絕對值最小則表明該實驗條件所對應的估計閾值最接近真實閾值。

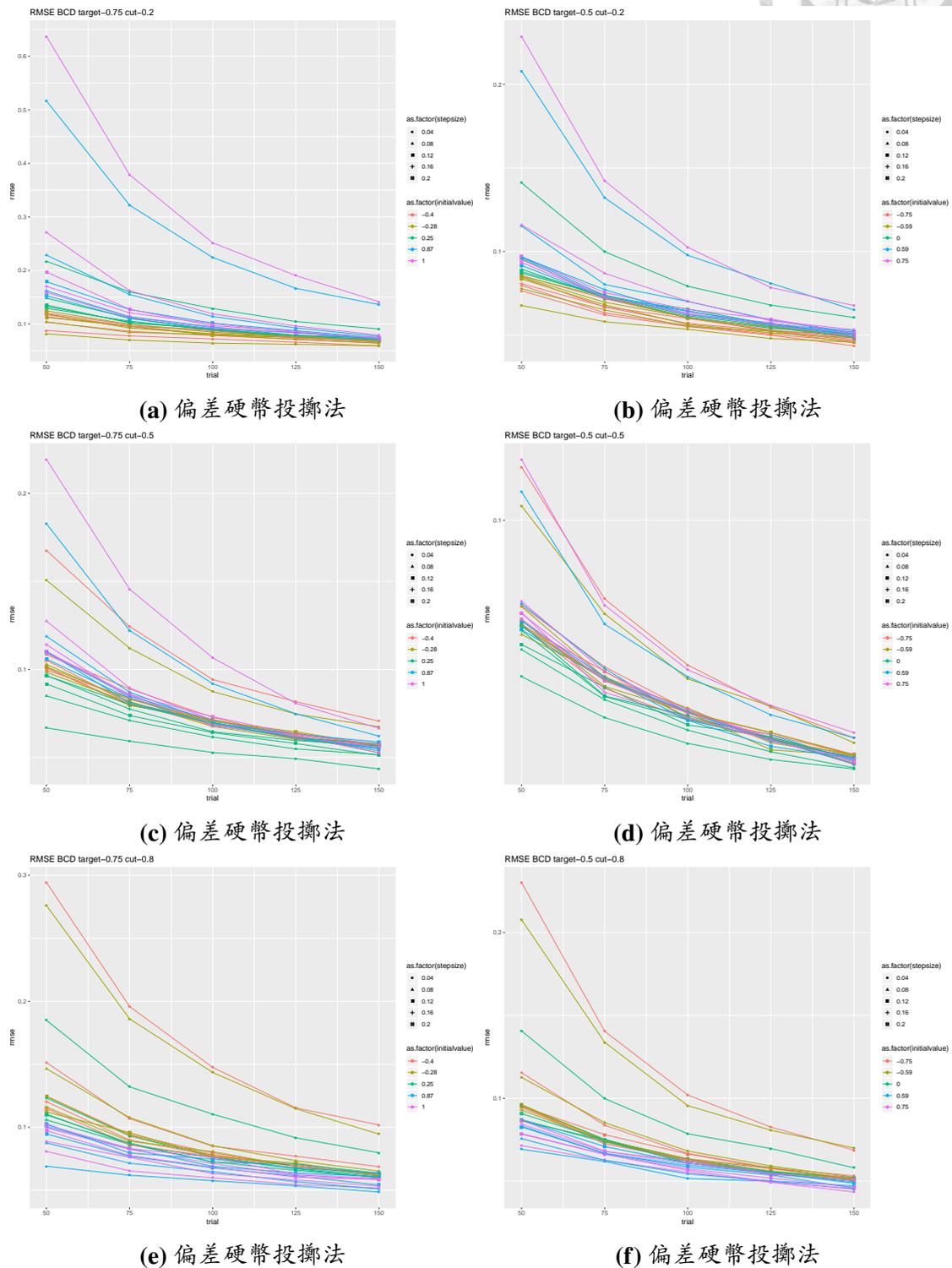
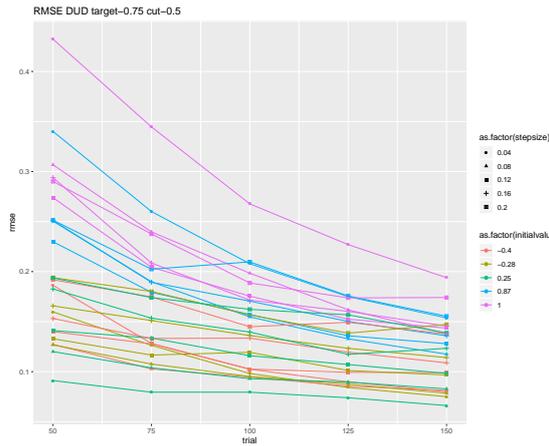
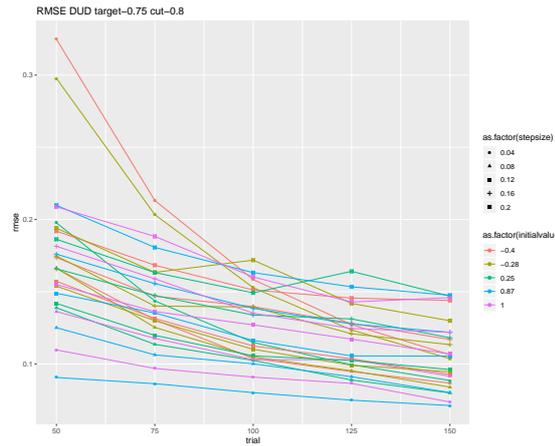


圖 8：偏差硬幣投擲法-均方根誤差

註：目標概率分別為0.75與0.5的均方根誤差結果，其適測方法為偏差硬幣投擲法。圖中縱坐標軸標記刻度是0.1但縱軸的範圍均不一致。



(a) 德爾曼上下法

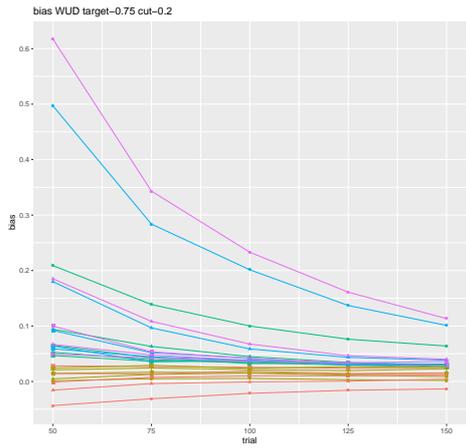


(b) 德爾曼上下法

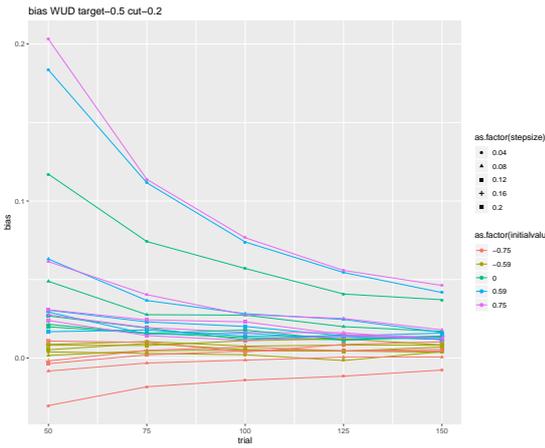
圖 9：德爾曼上下法-均方根誤差

註：目標概率為0.75，截點為0.5與0.8的均方根誤差結果，其適測方法為德爾曼上下法。圖中縱坐標軸標記刻度是0.1但縱軸的範圍與前述圖不一致。

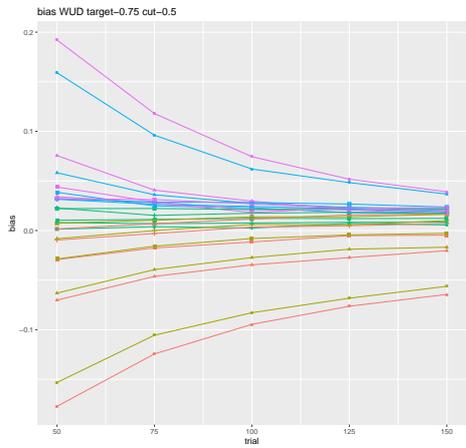
現對比三種適測方法的整體表現（見附錄一）：在目標概率為0.75的同等實驗條件下，相較於德爾曼上下法、偏差硬幣投擲法，帶權重上下法具有較小的均方根誤差（意即，同時擁有較小的最小值與較小的最大值），此表明帶權重上下法更有效率；而偏差硬幣投擲法與帶權重上下法則具有相似的偏誤量級，表明它們估計的準確度相當且優於德爾曼上下法。而在目標概率為0.5時，三種適測方法在均方根誤差與偏誤上差距不大，因此它們在引入回答信心量尺的閾值估計實驗裡具有彼此匹配的效率與準確度。



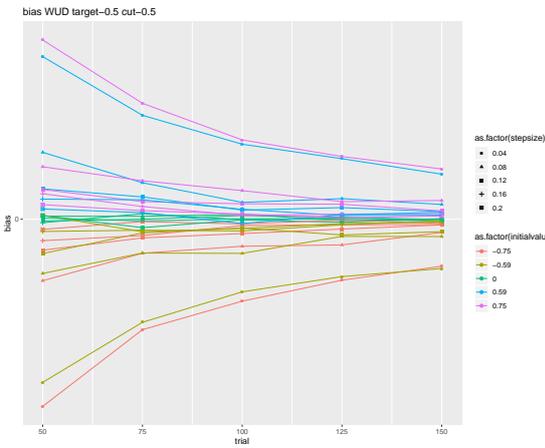
(a) 帶權重上下法



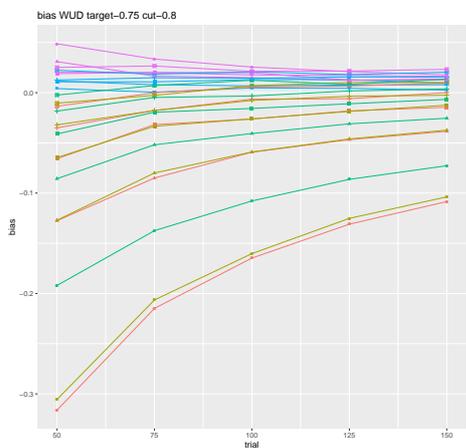
(b) 帶權重上下法



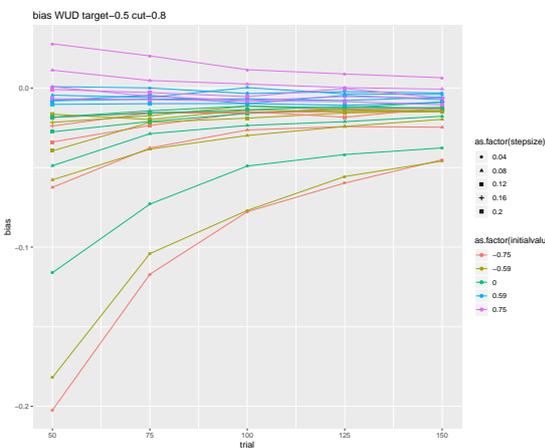
(c) 帶權重上下法



(d) 帶權重上下法



(e) 帶權重上下法



(f) 帶權重上下法

圖 10：帶權重上下法-偏誤

註：目標概率分別為0.75與0.5的偏誤結果，其適測方法為帶權重上下法。圖中縱坐標軸標記刻度是0.1但縱軸的範圍均不一致。



第五章 擴展

第一節 「同」/「異」作業與其非單調心理計量函數

與「是」/「否」作業不同，「同」/「異」作業的心理計量函數不具有單調性（見圖 11(a)）；在「同」/「異」作業中，當左右兩個刺激的強度差較大，此時受試者認為左右刺激一樣的概率較低，回答「是」（即「左邊刺激與右邊刺激一樣」）的概率亦相應較小（見圖 11(a)的兩端部分）；若兩個刺激的強度差減小至接近零時，受試者認為左右刺激一樣的概率增大，回答「是」的概率隨之增加。因此，在左右刺激的強度差由大轉小時（即圖 11(a)刺激強度由兩端向零點趨近），受試者回答「是」的概率相應上升（見圖 11(a)）。

若上述實驗的問題由「左邊刺激是否與右邊刺激一樣」改為「左邊刺激是否與右邊刺激不同」，可知在同樣的刺激強度（刺激強度與前文定義一致，即刺激強度=比較刺激-標準刺激）下，受試者回答「是」與原問題裡回答「是」為互補事件且這兩事件所對應的概率總和為1。因此，從原心理計量函數（見圖 11(a)）可描繪對應的變更問題後的心理計量函數（見圖 11(b)），本「同」/「異」模擬作業以變更後的實驗設計方式為基礎，即受試者應回答實驗問題「左邊刺激是否與右邊刺激不同」。

現對比「同」/「異」作業與「是」/「否」作業（見緒論）情境，可知其差異性：隨刺激強度由負向到正向變化（即從比較刺激小於標準刺激到比較刺激大於標準刺激的過程），「是」/「否」作業所對應的心理計量函數值（即回答「是」的概率）隨之增加；而「同」/「異」作業的心理計量函數值先減小後增

大（見圖 11(b)），因此該函數不再具有單調性。

在「是」/「否」作業裡，本文假設受試者回答信心的期望值與刺激強度成正相關，又因為回答信心的期望值反映的是受試者在原二回答選項中給出肯定回答的概率，即它與心理計量函數值亦成正相關。為保持兩個實驗作業裡模擬實驗設定的一致性：在「同」/「異」作業裡亦假設受試者回答信心的期望值與心理計量函數值（見圖 11(b)）具有正相關關係；因此受試者回答信心的期望值隨刺激強度由負向到正向的轉變應先減小後增大。同樣地，在刺激強度為零時，其對應的回答信心的期望值設在回答信心量尺的中點。

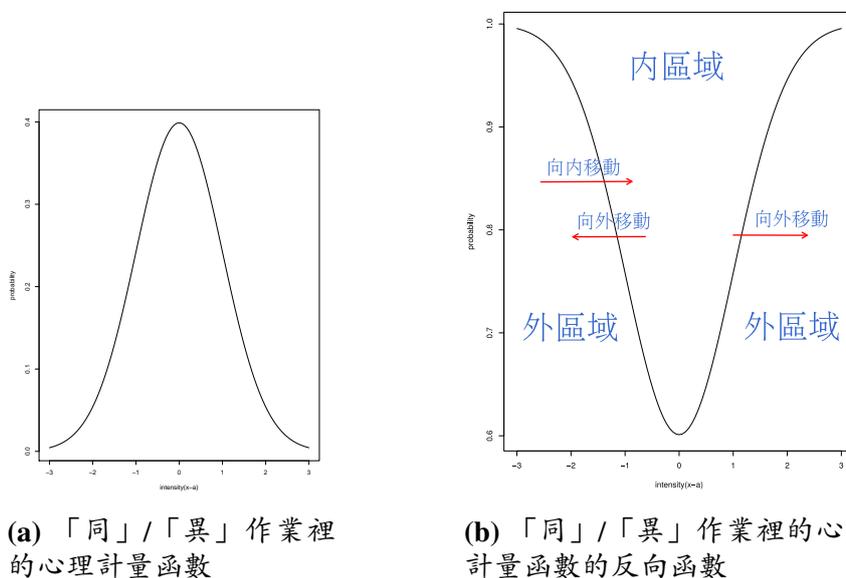


圖 11：「同」/「異」作業裡的心理計量函數

註：圖中橫坐標均表示刺激強度，左圖縱坐標代表受試者回答「是」的概率（即左邊刺激與右邊刺激一樣大）；右圖縱坐標代表受試者回答「是」的概率（即左邊刺激與右邊刺激不同）。兩圖在相同刺激強度下所對應的函數值和為1。

另一方面，當刺激強度為零時，受試者所感受到的區辨難度最大：受試者既難判斷「左邊刺激是否比右邊刺激大」（「是」/「否」作業）亦難判斷「左邊刺激是否與右邊刺激不同」（「同」/「異」作業），因此受試者在該條件下對其回答的不確定程度最高。這種高度不確定亦影響受試者每次在相同刺激強度下的回答信心，使其變化範圍較大。當刺激強度從零向兩邊推進時（即向正無窮與負無窮移動），左右刺激差值變大，受試者所感受到的區辨難度減小，相應地對

其回答的確定程度增加，受試者每次在該刺激強度下的回答信心變化幅度很小。

綜上，本模擬實驗設定在某一固定刺激強度下，其回答信心分布的期望值與變化範圍由以下等式（等式(14)與等式(15)）決定。其中， i 是刺激強度， $dnorm(i, mean, std)$ 是均值為 $mean$ ，標準差為 std 的常態分布函數的密度函數，而 u 是等式(14)的左邊。由等式(14)可知，在刺激強度為零時，回答信心的期望值設定在回答信心量尺的中點。當刺激強度的絕對值增大時，其回答信心的期望值相應增大而回答信心的變化範圍則減小（見等式(14)與等式(15)）。

$$mean\ of\ confidence = 1 - 0.5 / dnorm(0, 0, 0.5) * dnorm(i, 0, 0.5) \quad (14)$$

$$standard\ deviation\ of\ response\ confidence = \sqrt{0.25u(1 - u)} \quad (15)$$

儘管模擬實驗參數已設定完畢，但本文採用的三種適測方法均要求心理計量函數具有單調遞增性，而「同」/「異」作業的心理計量函數並不滿足該特性，因此，在非單調心理計量函數中採用適測方法需稍加處理。本文主要參考Hsu與Doble（2015）的研究解決此問題：他們提出轉換心理計量函數，使其滿足函數的單調遞增性，再結合傳統適測方法進行實驗。此外，García-Pérez（2014）亦提出另一解決方式：他選擇保持原心理計量函數形態並提出新的適測方法做實驗。以下分別陳述這兩種做法，並探討它們的異同點。

第二節 非單調心理計量函數中適測方法的應用

傳統的適測方法對心理計量函數單調性的要求造成它無法適用於非單調的心理計量函數，為解決此問題，Hsu與Doble（2015）提出先轉換非單調的心理計量函數使其滿足單調性，在轉換後的心理計量函數中採用傳統的適測方法即可。本模擬實驗以此方法為基礎，對其適當變形以適用於引入回答信心量尺的

閾值估計實驗，因此「同」/「異」作業的心理計量函數經轉換後的心理計量函數如下（見等式(16)，該心理計量函數推演細節見附錄三）。其中，*cutoff*是截點， $F_{x-a<0}$ 是比較刺激 x 小於標準刺激 a 時回答信心分布所對應的累積分布函數，其均值為 u （由等式(14)計算所得，等式(14)中刺激強度 $i = x - a$ ），標準差為 σ （由等式(15)計算所得）；同理， $F_{x-a>0}$ 是比較刺激大於標準刺激時回答信心函數所對應的累積分布函數。由於轉換後的心理計量函數（見等式(16)）是單調遞增函數，在此基礎上採用前述三種適測方法（見第二章「適測方法」）可估計引入回答信心量尺的「同」/「異」作業的閾值。

$$\begin{aligned}
 P_{cutoff}(x - a) &= F_{x-a<0}(1 - cutoff, u, \sigma) \quad \text{when } x < a, \\
 &= 1 - F_{x-a>0}(cutoff, u, \sigma) \quad \text{when } x \geq a
 \end{aligned}
 \tag{16}$$

同樣基於對適測方法在非單調心理計量函數中適用性的考慮，García-Pérez（2014）設計新的適測方法使其適用於非單調心理計量函數，以此保持心理計量函數的原有形態。該做法將心理計量函數（見圖 11(b)，以此為例）分為「外區域」與「內區域」，內區域是由函數與上漸近線圍成的封閉區域，除此之外的外圍部分是外區域。若刺激強度由內區域向外區域方向變化，則刺激強度「向外」調整；反之，刺激強度「向內」調整。該適測方法做法如下：若受試者給予肯定回答，則下次試驗的刺激強度向內移動；若受試者給予否定回答，則下次試驗的刺激強度「隨機向外」移動（在García-Pérez（2014）的研究中，「隨機向外」指刺激強度可能向外或向內移動，其移動方向由一柏努利隨機變量決定，該柏努利隨機變量取向內移動的概率為0.5，因此向外移動概率亦為0.5。）。

儘管本文採用Hsu與Doble（2015）的轉換方式與傳統適測方法相結合的做法模擬實驗，但García-Pérez（2014）的研究亦解決在非單調心理計量函數中使

用適測方法的問題。因此以下先探討這兩種做法所對應實驗目的差異，並比較García-Pérez (2014) 的適測方法與傳統適測方法的異同點。

從實驗可行性分析，由於在估計單調心理計量函數閾值的實驗裡，目標概率與閾值的對應關係具有唯一性，因此這種實驗可估計預定目標概率所對應的閾值且估計結果唯一。而估計非單調心理計量函數閾值（見圖 11(b)）存在一個問題：目標概率與閾值的對應關係不具有唯一性，其目標概率對應的閾值個數有三種情況：兩個、一個或不存在（實驗選定目標概率低於心理計量函數最低值）。因此，若研究者希望估計非單調心理計量函數閾值需解決此問題。

Hsu與Doble (2015) 通過轉換心理計量函數可避免此問題的出現：其轉換後的心理計量函數因滿足單調性，亦滿足目標概率與閾值的唯一對應關係。由於García-Pérez (2014) 的實驗目的是估計完整心理計量函數，且保留原心理計量函數形態不做轉換，因此目標概率與閾值不具有唯一對應關係，該研究方法無法估計閾值。

另一方面，若實驗目的是估計完整的心理計量函數：García-Pérez (2014) 的研究即以此為目標；Hsu與Doble (2015) 的方法與適測方法結合雖用於閾值估計實驗，但亦可能實現估計完整的心理計量函數。閾值估計實驗能估計目標概率所對應的閾值，將該閾值帶入心理計量函數所得即為目標概率，因此每個閾值與其對應的目標概率構成心理計量函數的一個數據點。若重復多次閾值實驗，每次實驗皆估計不同目標概率所對應的閾值，這些數據點可用於估計完整心理計量函數。

雖然兩種方法皆可估計心理計量函數，但García-Pérez (2014) 的做法需預先假設心理計量函數的形態（需具體假設該函數表達式），在「同」/「異」作業裡採用新的適測方法，記錄刺激強度與其對應的回答結果（即「是」或「否」），待試驗數達到上限後終止並以這些數據為基礎採用最大概似估計方法（maximum likelihood method）估計該心理計量函數的參數。該做法可能有一個

潛在問題：當心理計量函數的函數形態未知，若研究者假設的心理計量函數錯誤，則該適測方法所估計的函數參數無效，它所描繪的心理計量函數亦不正確。此外，因為判斷心理計量函數的假設合理性並不屬於適測方法的研究範圍，這種適測方法無法判斷心理計量函數的假設是否合理，亦難以從此角度評定實驗結果的合理性。

若採用Hsu與Doble（2015）的做法先將非單調心理計量函數轉換為單調心理計量函數，再採用傳統的適測方法估計不同閾值，多次實驗結果可繪出完整心理計量函數。但該做法無需假設心理計量函數的具體形態，因此可避免對未知心理計量函數做過多假設。

另一方面，從實驗效率而言，García-Pérez（2014）的適測方法估計完整心理計量函數效率較高（見García-Pérez（2014）圖4對 β 的估計）：該適測方法在200個試驗數以上能較精確估計心理計量函數的參數。若採用重復多次閾值估計實驗估計不同閾值以收集數據點的方式描繪心理計量函數則需更多的試驗數，以保證其最終估計閾值的準確度。

現分別探討三種適測方法與García-Pérez（2014）的適測方法相比的異同點。需在此說明的是，因本模擬實驗已採用Hsu與Doble（2015）將「同」/「異」作業的心理計量函數轉換成單調的心理計量函數，在此基礎上傳統的適測方法選取刺激強度的移動方式與「是」/「否」作業中的移動方式相同，即刺激強度以增加（上升）或減少（下降）的方式移動。而在García-Pérez（2014）的適測方法裡，刺激強度的移動方式以「向內」或「隨機向外」的方式移動。

García-Pérez（2014）的方法與帶權重上下法（見第二章「文獻回顧」·「帶權重上下法」）均以上下階梯法（見第二章「文獻回顧」·「適測方法」部分）為基礎，因而具有其相似性。在García-Pérez（2014）的方法裡，它規定下次試驗的刺激強度向內與「隨機向外」的步伐在整個實驗裡恆定不變，但向內與「隨

機向外」的步伐並不一定相等。比如，當實驗採取「1外3內」的刺激選取法則（以圖 11(b)的心理計量函數為例）：若受試者給予肯定回答，下次試驗的刺激強度向內移動三分之一步長；反之，下次試驗的刺激強度「隨機向外」（方向隨機）移動一個步長。類似地，在帶權重上下法裡，刺激強度每次上升與下降的步伐不變但不一定相等（刺激強度的調整步伐由迭代式決定）。

另一方面，帶權重上下法沒有引入任何的隨機變量：在「同」/「異」作業裡，受試者給予肯定回答（即左右刺激不同）後下次試驗刺激強度必然減小；反之，受試者給予否定回答必增加下次試驗刺激強度。而在García-Pérez（2014）的適測方法裡，當受試者給予否定回答，此時刺激強度「隨機向外」移動，它的「隨機向外」移動方向存在兩種情況：即刺激強度可能向內移動亦可能向外移動（見圖 11(b)移動方向）；因此該方法引入柏努利隨機變量決定每次「隨機向外」移動時，刺激強度具體應向外或向內移動（見圖 11(b)）。

這種引入柏努利隨機變量的做法，在García-Pérez（2014）的適測方法裡可能導致一種情境的產生：當受試者給予否定回答時，表明他已無法區辨左右刺激的差異²（見圖 11(b)），但刺激強度仍可能向內移動，即向左右刺激強度差值更小的方向移動。但從它與帶權重上下法的基礎——上下階梯法的觀點出發，當受試者能區辨左右刺激的差異時應減小兩個刺激的強度差值；反之，若受試者已無法區辨兩個刺激的差異，則應增加兩個刺激的強度差值，使其差異更為明顯。因此，García-Pérez（2014）的適測方法在引入柏努利隨機變量後不完全滿足上下階梯法的法則，而帶權重上下法仍滿足此法則。

與之相對，在德爾曼上下法裡（見第二章「文獻回顧」·「德爾曼上下法」，此處取目標概率高於0.5的情境），當受試者給予肯定回答，表明受試者認為左右刺激不同，下次試驗刺激強度仍可能增加一個步長，即向左右刺激強度差更大的方向移動。因此，這一做法與德爾曼上下法更類似：兩者都引入柏努利

²因為研究問題是「左邊刺激是否與右邊刺激不同」，受試者給予否定回答，表明受試者傾向於認為左右刺激一樣。

隨機變量且下次試驗的刺激強度一定發生改變。更進一步地，儘管偏差硬幣投擲法（見第二章「文獻回顧」·「偏差硬幣投擲法」）亦引入柏努利隨機變量，但根據其迭代式可知，當受試者給予肯定回答時，其下次試驗刺激強度可能維持不變。

但García-Pérez（2014）引入的柏努利隨機變量與另外兩個適測方法不同：在德爾曼上下法與偏差硬幣投擲法裡，它們的柏努利隨機變量的概率隨目標概率的位置變化；García-Pérez（2014）的適測方法定義柏努利隨機變量的概率恆定為0.5。這種差異由實驗目的所決定，德爾曼上下法與偏差硬幣投擲法均用於閾值估計，其適測方法的迭代式需收斂於閾值；而García-Pérez（2014）的適測方法是估計完整的心理計量函數，因此，後者並不需要趨近固定的點。

綜上，García-Pérez（2014）的適測方法因其以估計心理計量函數為目標，使其與傳統的適測方法有差異；且該實驗目的也導致這種適測方法需對心理計量函數有較具體嚴格的假設，這要求研究者預先對實驗的心理計量函數有一定的了解。若心理計量函數的形態未知，研究者需先進行相關實驗獲得心理計量函數的形態，再應用該適測方法進行實驗。雖然該適測方法有這種侷限性，但充分假設帶來的優勢是它可用較少的試驗數估計完整心理計量函數。

另一方面，若研究者只需研究心理計量函數的某個點，研究者僅需估計閾值而無需完整心理計量函數；採用García-Pérez（2014）的適測方法可能耗費不必要資源，且研究者需預先確定心理計量函數形態。

若先採用Hsu與Doble（2015）的方式將「同」/「異」作業的心理計量函數轉換為單調的心理計量函數，再使用傳統的適測方法，不僅可用於閾值估計實驗，亦可比較各種適測方法在「同」/「異」作業的表現。且該轉換無需修改原先的適測方法，亦無需引入嚴格的心理計量函數假設，因此擴展了它的適用範圍。

第三節 適測方法、截點、目標概率、步長、初始值、試驗數、模

擬次數



本模擬實驗的適測方法、截點、目標概率、步長、試驗數、模擬次數均與「是」/「否」作業一致（見第三章「模擬實驗」）故此不再贅述。但需在此指出的是，本文指定「同」/「異」作業裡的目標概率是指心理計量函數轉換後（即單調心理計量函數）的目標概率；此目標概率為0.75時，其對應於原非單調心理計量函數中所對應目標概率會低於原心理計量函數目標概率的0.75。目標概率為0.75時，初始值分別為-0.65, -0.49, 0.46, 0.97, 1.1（-0.49, 0.46, 0.97分別在截點0.2, 0.5, 0.8對應的閾值附近）；在目標概率為0.5時，初始值分別為-0.9, -0.76, 0, 0.76, 0.85（-0.76, 0, 0.76分別在截點0.2, 0.5, 0.8對應的閾值附近）。均方根誤差與偏誤按第三章「均方根誤差與偏誤」計算。

第四節 結果

壹、估計目標概率圖

偏差硬幣投擲法在「同」/「異」作業的實驗裡（見圖 12(a)(b)(e)(f)），估計目標概率的偏差方向嚴重受截點位置影響：低截點導致過高估計目標概率；高截點引起整體低估目標概率。在小步長（步長為0.04或0.08）條件下，隨初始值偏離閾值越遠，其估計目標概率的偏差越大。當步長增大時，初始值對估計概率點的影響減小，不同初始值所對應的估計點越接近。在此條件下截點位置與目標概率共同決定估計點的重合度：目標概率為0.5時（見圖 12(a)(c)(e)），步

長較大時估計點均重合，表明初始值對估計目標概率的影響近乎消失；而目標概率為0.75時（見圖 12(b)(d)(f)），在低截點與中間截點時，即使在最大步長的條件下，初始值仍影響其最終估計點的位置，因而導致估計點無法重合。因此在「同」/「異」作業裡，初始值對該適測方法的影響無法完全通過增大步長抵消。

另一方面，適測方法在中間截點的估計結果可反映在不受截點位置的影響下，它估計目標概率的偏差情況。因此，在中間截點位置（見圖 12(c)(d)）、目標概率為0.75時，隨步長增大偏差硬幣投擲法低估目標概率的程度越嚴重，表明該適測方法在該實驗情境裡更易低估目標概率；而目標概率為0.5時則不具有這種傾向性。

在目標概率為0.5條件下，德爾曼上下法與帶權重上下法表現均與偏差硬幣投擲法類似：在低截點時整體傾向高估目標概率，在高截點時則低估目標概率。在步長較大且截點居中時，它們均能準確估計目標概率。但在目標概率為0.75條件下（見圖 13），它們卻與偏差硬幣投擲法表現有所不同。

在中間截點時，對比德爾曼上下法與偏差硬幣投擲法（見圖 12(d)與圖 13(a)）可知，偏差硬幣投擲法隨步長增大呈現低估的趨勢；而德爾曼上下法雖受初始值影響，整體卻趨近於目標概率。帶權重上下法亦與德爾曼上下法有相似結果（見圖 13(a)與圖 13(c)）。在高截點條件下，隨步長增大德爾曼上下法整體呈現高估目標概率的趨勢（見圖 13(b)）；而偏差硬幣投擲法則整體呈現低估目標概率的趨勢（圖 12(f)）。相較於兩者，隨步長增大帶權重上下法則穩定地收斂於目標概率（圖 13(d)）。

綜上，目標概率為0.5時，隨步長增大，初始值對三種適測方法的影響逐漸消失。在目標概率為0.75時，僅在高截點條件下，增大步長可抵消初始值對適測方法的影響。若對比同一適測方法在兩個目標概率（0.5與0.75）的模擬結果：目標概率為0.5時，三種適測方法的表現均更好，此表明目標概率的位置亦影響適

測方法的精確度。此外，本文在此僅展示試驗數100時的模擬結果，但試驗數越少上述結果越顯著：即已表現出估計偏差的實驗條件會造成更大的估計偏差；若試驗數增加，初始值的影響則會相應減小。



貳、均方根誤差

由偏差硬幣投擲法的均方根誤差結果可知（見圖 14），除目標概率為0.5且位於中間截點的實驗條件（見圖 14(d)），其餘實驗條件結果具有如下特徵：最小步長時，偏離閾值最遠的初始值擁有最大均方根誤差，最接近閾值的初始值有最小的均方根誤差。當步長增大時，因初始值的影響減小，不同步長所對應的均方根誤差折線隨試驗數的增加而開始重合。隨試驗數增加，均方根誤差折線的下降速率減緩，表明試驗數的增加對減小誤差的作用越來越小。

此外，目標概率為0.75所對應均方根誤差折線的下降幅度大於同等條件下目標概率為0.5所對應的均方根誤差折線的下降幅度（分別比較圖 14(a)(b)、圖 14(c)(d)、圖 14(e)(f)），因此，在目標概率為0.5時，試驗數的增加對其實驗結果的提高程度小於對目標概率為0.75時實驗結果的提高。

但在目標概率為0.5且截點在中間位置的實驗條件下（見圖 14(d)），最小的步長與最合適的初始值（即初始值位於閾值附近）僅在試驗數較小時具有優勢，當試驗數逐漸增大時，其他實驗條件所對應的均方根誤差迅速減小。儘管該結果看似與其他目標概率與截點位置組合的實驗條件結果不同；仔細觀察該結果發現隨試驗數增大，整體均方根誤差折線雖呈現下降趨勢，仍有上下起伏的波浪狀。又因為初始值為零時，離閾值非常接近，因而當步長太小時，適測方法每次選取的刺激會在閾值附近滑動，雖不會偏離閾值太遠，卻也無法滑動到閾值。若步長較大，由於擺動較大，在試驗數增大時反而更大概率滑動到閾值。因此，看

似最佳的實驗條件（即最小步長與最接近閾值的初始值組合）在此情境中反而表現更差。小步長的劣勢亦體現在目標概率為0.75時，低截點的實驗條件（見圖 14(a)）：其最高的若干條均方根誤差折線均為最小步長的實驗條件，表明在「同」/「異」作業裡步長過小導致均方根誤差下降緩慢，因此在最大試驗數條件下仍擁有過高的均方根誤差。

另一方面，在目標概率為0.75時（見附錄二），對比各個截點的最大均方根誤差可知：低截點條件有最大的均方根誤差，高截點時則有最小的均方根誤差。同樣地，對比各個截點的最小均方根誤差（見附錄二），低截點的均方根誤差最大，高截點仍具有最小的均方根誤差。而目標概率為0.5時（見附錄二）則中間截點所對應的整體均方根誤差變化幅度較小；其他兩個截點位置所對應的均方根誤差整體差距不大，它們的均方根誤差亦差距較小。

德爾曼上下法與偏差硬幣投擲法在一定程度上具有其相似性（尤其在目標概率為0.5的條件裡），故不再贅述。本文在此僅探討其差異性。在目標概率為0.75的低截點條件下（見圖 15(a)），德爾曼上下法的均方根誤差折線分布鬆散；但該圖底部最小均方根誤差折線組對應不同步長下，初始值均為-0.65或-0.49（即最接近閾值的初始值）的實驗條件。而最大的均方根誤差折線組（圖 15(a)上方折線組）則對應最小的步長與最遠的初始值，且該折線組在最大試驗數時仍具有頗為顯著的高均方根誤差。該結果表明德爾曼上下法受初始值影響大於其所受步長影響：離閾值較近的初始值比增大步長更加有效地降低誤差；若初始值偏離閾值過遠，增大步長對減小誤差的效果極為有限。

另一方面，德爾曼上下法對初始值的依賴性隨截點位置的升高而減小（見圖 15(b)(c)）：在高截點條件下，最小的步長與最優初始值仍具有最小均方根誤差；隨試驗數增大，距離閾值最遠的初始值即使在小步長條件下，其誤差仍快速下降。該結果表明德爾曼上下法在高截點位置時受初始值的影響減小，且提高試驗數能有效降低不合適的初始值造成的誤差。又因為目標概率為0.5時，德爾曼

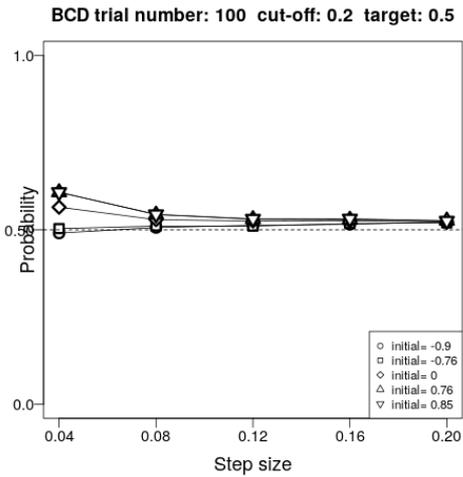
上下法在不同截點位置不具有這種差異。因此，德爾曼上下法在不同截點位置，對初始值的敏感度差異亦受目標概率位置影響；且增大試驗數不一定能有效降低誤差水平。

綜上，德爾曼上下法對初始值的依賴程度表明，即使增大步長與試驗數亦不能抵消初始值影響；因此在低截點條件下，德爾曼上下法的效率比偏差硬幣投擲法低（即收斂速度過慢）。而帶權重上下法與偏差硬幣投擲法表現基本無異，故此不再贅述。此外，比較各個適測方法均方根誤差在各組實驗條件下的最值（見附錄二）可知，在目標概率為0.75的條件下，偏差硬幣投擲法與帶權重上下法表現稍優於德爾曼上下法；在目標概率為0.5時三種適測方法無明顯差距。

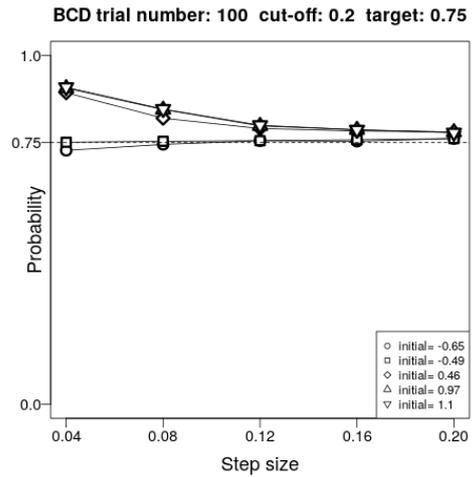
參、偏誤

本文在此以帶權重上下法為主說明偏誤結果。在帶權重上下法的偏誤結果（見圖 16）可知，當初始值大於閾值時，多數實驗條件所對應的偏誤為正，即估計閾值高於閾值本身；反之，初始值小於閾值時，多數實驗條件所對應的偏誤傾向估計閾值小於閾值。而在目標概率為0.75的條件下（見圖 16(a)(c)(e)），低截點位置所對應偏誤的浮動範圍遠高於另兩個截點位置所對應的偏誤浮動範圍；但在目標概率為0.5的條件下不具有這種特徵（見圖 16(b)(d)(f)）。

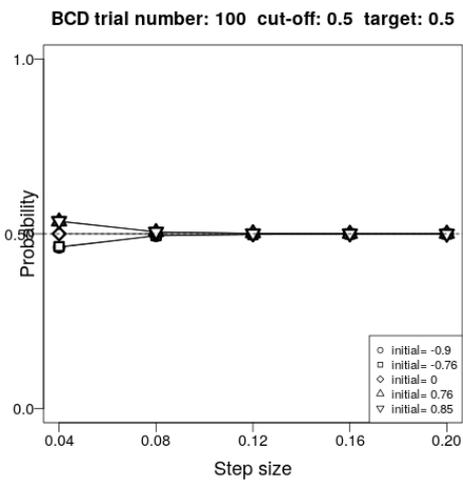
另一方面，在目標概率為0.5時（見圖 16(b)(d)(f)），各組實驗條件所對應的偏誤折線重疊度高於同樣截點位置下目標概率為0.75的偏誤結果（見圖 16(a)(c)(e)），表明在目標概率為0.5時，帶權重上下法有較小的偏誤且偏誤的分散程度較小（亦可見附錄二偏誤最值）。因此該適測方法在目標概率為0.5的表現比同等條件下在目標概率為0.75時表現更好，亦表明帶權重上下法的模擬結果受目標概率位置所影響。



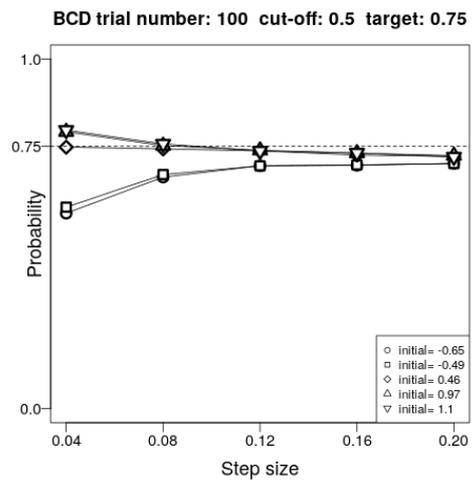
(a) 偏差硬幣投擲法



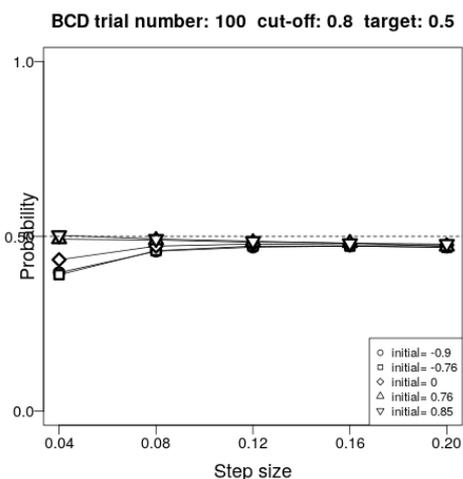
(b) 偏差硬幣投擲法



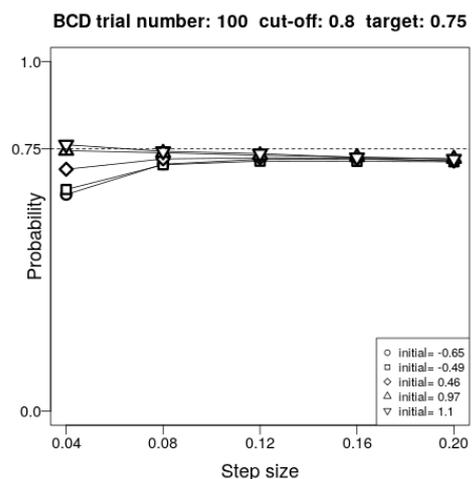
(c) 偏差硬幣投擲法



(d) 偏差硬幣投擲法



(e) 偏差硬幣投擲法



(f) 偏差硬幣投擲法

圖 12：偏差硬幣投擲法-估計目標概率圖

註：目標概率分別為0.75與0.5時，各截點所對應估計目標概率圖，其適測方法為偏差硬幣投擲法。

此外，對比不同步長的實驗結果（見圖 16）可知，當步長較大時，增加試驗數對減小偏誤的效果很小；步長較小時，增加試驗數能有效減小偏誤。因為步長較大時，實驗的刺激強度已快速收斂到閾值附近，此時增加試驗數無法顯著改善試驗結果；在步長較小的條件下，因其收斂速度過慢，增加試驗數能確保其最後收斂於閾值附近，從而減小偏誤。

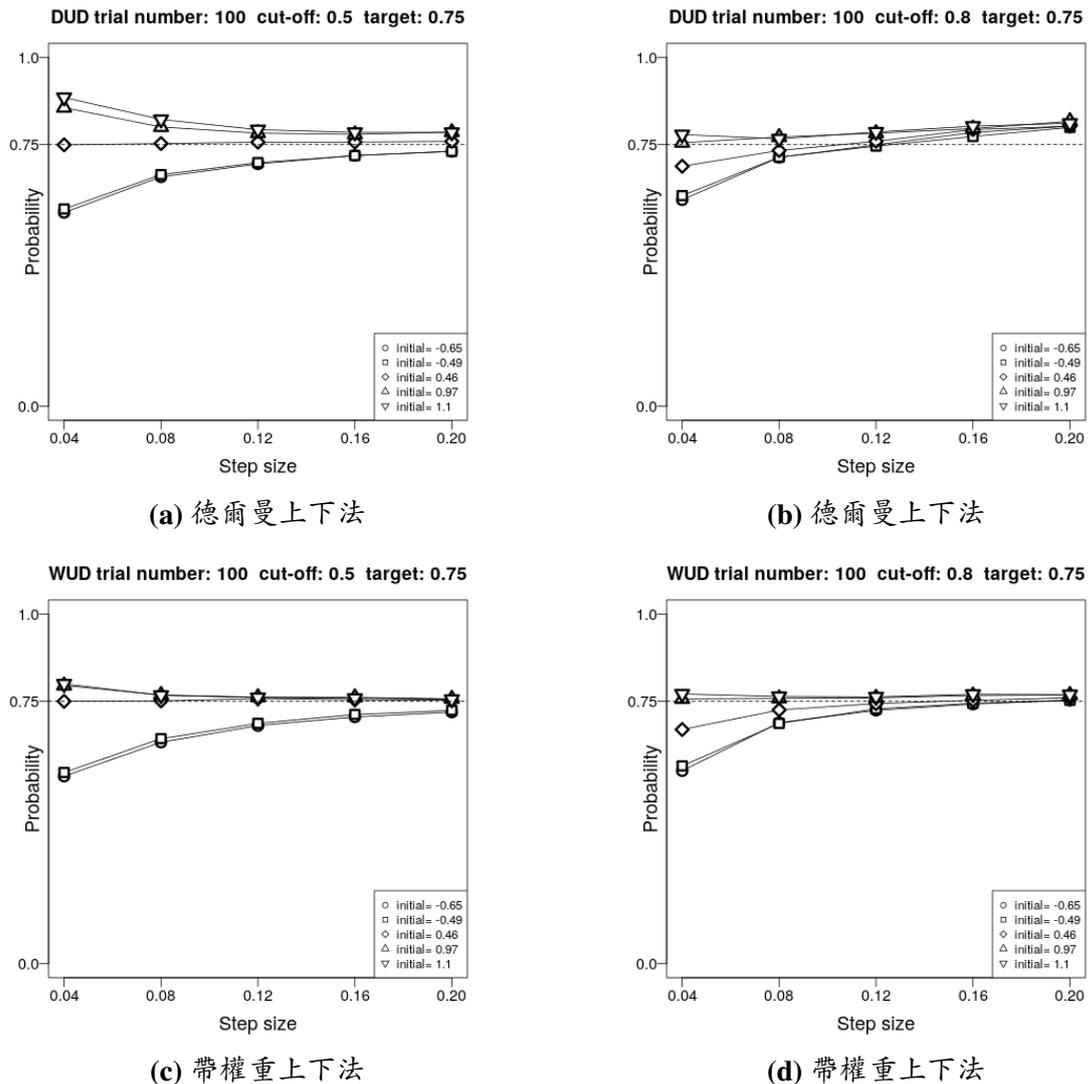
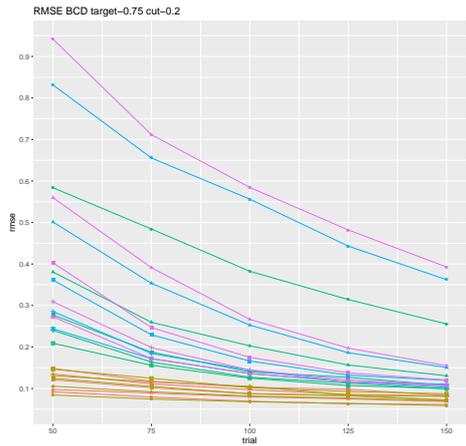


圖 13：德爾曼上下法與帶權重上下法-估計目標概率圖

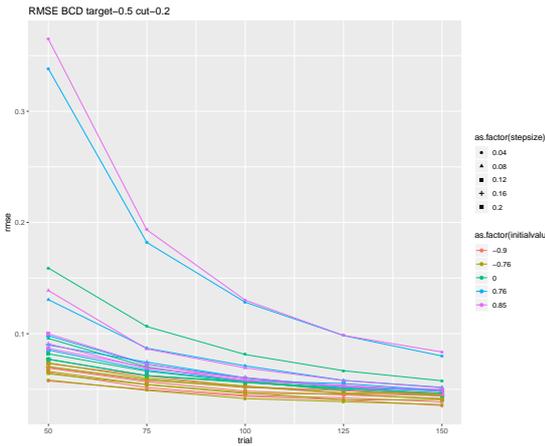
註：目標概率為0.75時，中間截點與高截點對應的估計目標概率圖，其適測方法為德爾曼上下法與帶權重上下法。

由於另兩種適測方法的偏誤結果與帶權重上下法相近，故不在此贅述其相似性，僅表述其差異性。整體比較各個適測方法偏誤在各組實驗條件下的最值（見附錄二）可知，德爾曼上下法具有最大與最小的偏誤，表明德爾曼上下法的偏誤

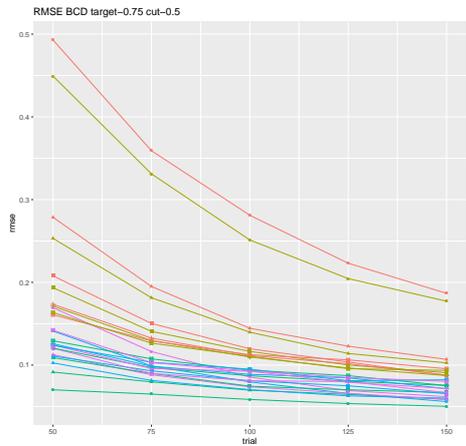
散度更大；而其他兩種適測方法的偏誤有較穩定的上下限。在目標概率為0.75、高截點的實驗條件下（見圖 17(a)(b)與圖 16(e)），以偏誤為零的結果分界發現：相比於帶權重上下法的偏誤折線圍繞於零線的兩邊，偏差硬幣投擲法幾乎整體偏誤折線均位於零線下方，德爾曼上下法則大體位於零線上方。該結果表明在該實驗條件下，偏差硬幣投擲法整體有低估閾值的傾向，因而亦低估目標概率；德爾曼上下法則有高估閾值的傾向，進而導致整體亦高估目標概率。該結果亦與目標概率估計圖的模擬結果相符（偏差硬幣投擲法見圖 12(f)、德爾曼上下法見圖 13(b)）。



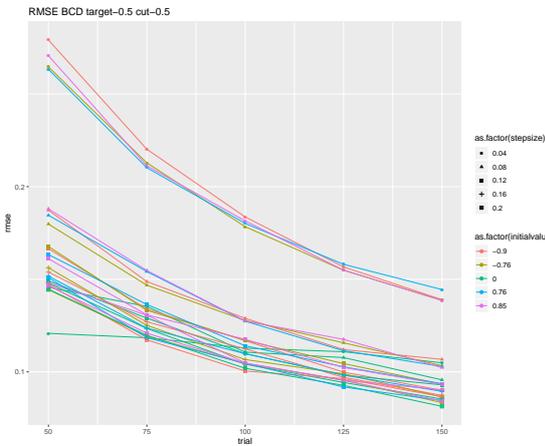
(a) 偏差硬幣投擲法



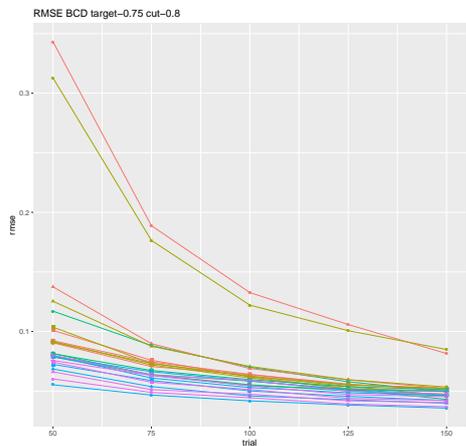
(b) 偏差硬幣投擲法



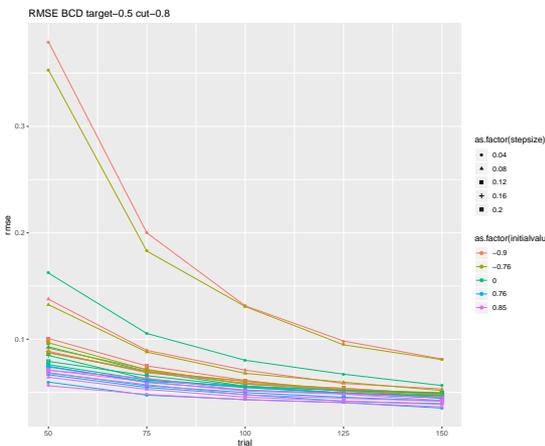
(c) 偏差硬幣投擲法



(d) 偏差硬幣投擲法



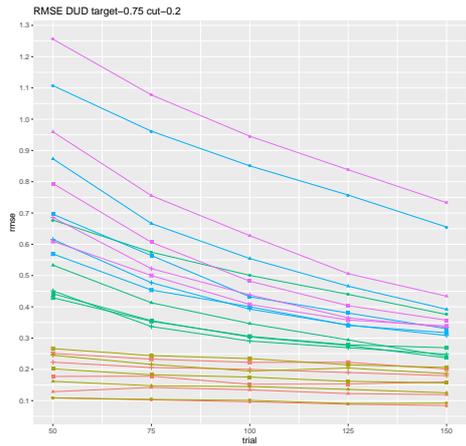
(e) 偏差硬幣投擲法



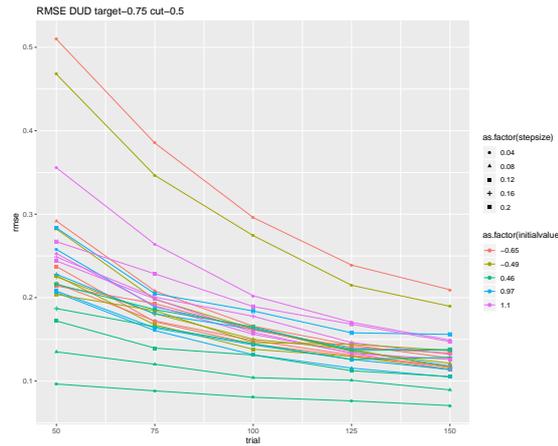
(f) 偏差硬幣投擲法

圖 14：偏差硬幣投擲法-均方根誤差

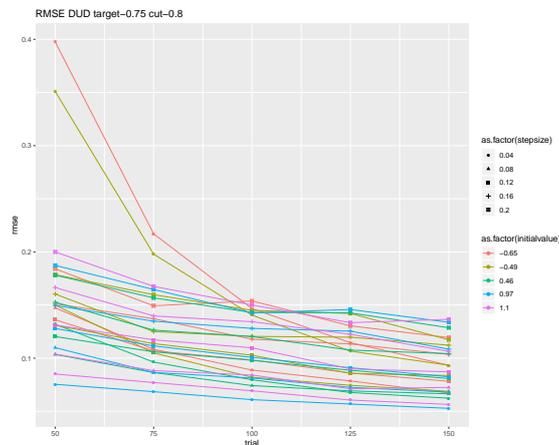
註：目標概率分別為0.75與0.5的均方根誤差結果，其適測方法為偏差硬幣投擲法。圖中縱坐標軸標記刻度是0.1但縱軸的範圍均不一致。



(a) 德爾曼上下法



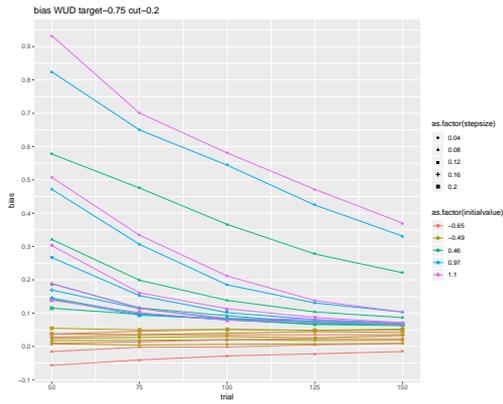
(b) 德爾曼上下法



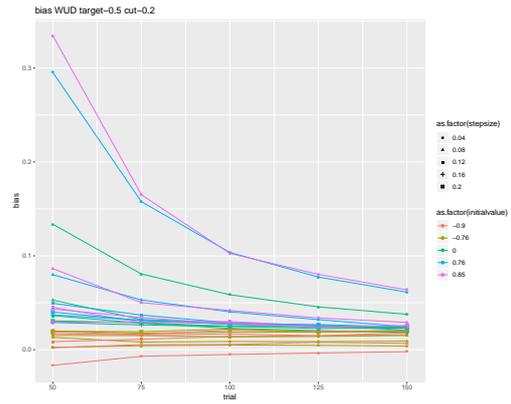
(c) 德爾曼上下法

圖 15：德爾曼上下法-均方根誤差

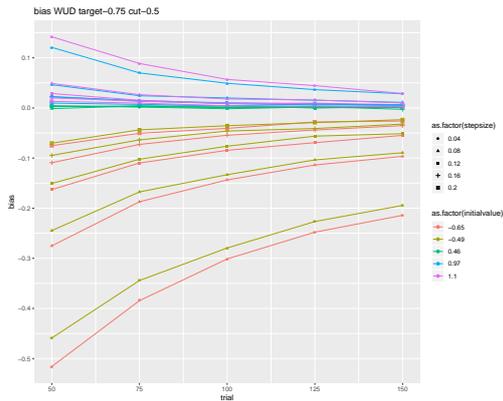
註：目標概率分別為0.75的均方根誤差結果，其適測方法為德爾曼上下法。圖中縱坐標軸標記刻度是0.1但縱軸的範圍均不一致。



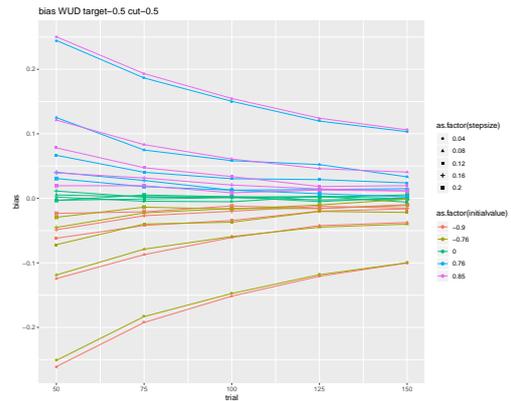
(a) 帶權重上下法



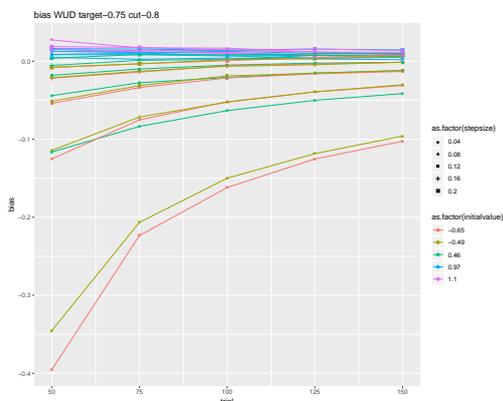
(b) 帶權重上下法



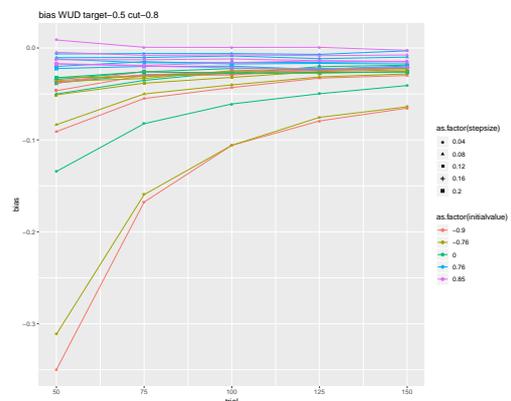
(c) 帶權重上下法



(d) 帶權重上下法



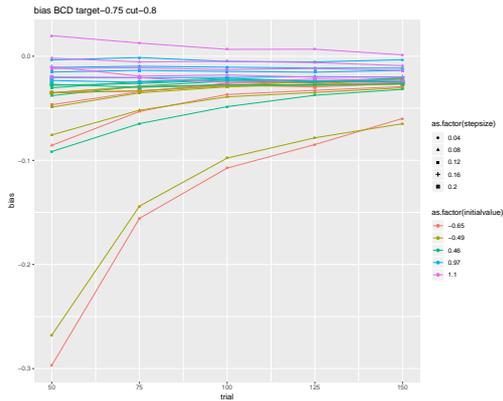
(e) 帶權重上下法



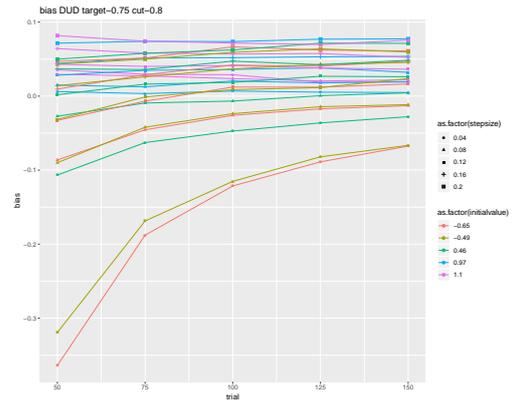
(f) 帶權重上下法

圖 16：帶權重上下法-偏誤

註：目標概率分別為0.75與0.5的偏誤結果，其適測方法為帶權重上下法。圖中縱坐標軸標記刻度是0.1但縱軸的範圍均不一致。



(a) 偏差硬幣投擲法



(b) 德爾曼上下法

圖 17：偏差硬幣投擲法、德爾曼上下法-偏誤

註：目標概率分別為0.75，高截點的偏誤結果，其適測方法為偏差硬幣投擲法、德爾曼上下法。

圖中縱坐標軸標記刻度是0.1但縱軸的範圍均不一致。



第六章 總結

本文探討在「是」/「否」作業與「同」/「異」作業裡引入回答信心量尺對偏差硬幣投擲法、德爾曼上下法，帶權重上下法估計閾值的影響，通過電腦模擬實驗分析各個因素作用下三種適測方法的估計結果。傳統閾值估計實驗受實驗設計所限制，研究者無法區別受試者在確定與不確定情境下給予的回答，亦無法根據實驗需求剔除受試者盲猜的回答。為蒐集受試者對其回答的信心程度這一資訊，本文引入回答信心量尺蒐集受試者每次回答的信心程度並根據預先選定的信心標準切割受試者的回答為「是」或「否」，並以此為出發點分析該量尺在閾值估計實驗裡對不同適測方法的影響。

在「是」/「否」作業裡，本模擬結果顯示回答信心的截點標準與目標概率均會影響三種適測方法使其偏離閾值，但增大步長與合適的初始值可減小偏差。而不同步長與初始值的組合隨試驗數的增加，均方根誤差逐漸減小，這表明增加試驗數在多種條件下均可改進適測方法的表現。此外，對比三種適測方法可知，在這類實驗裡德爾曼上下法更適合採用較小步長，而偏差硬幣投擲法與帶權重上下法更適合選取較大步長。且步長越大適測方法收斂到目標區域的速度越快，這表明偏差硬幣投擲法與帶權重上下法比德爾曼上下法更有效率。同時，偏差硬幣投擲法與帶權重上下法的偏誤與均方根誤差較小，因此在這類作業裡偏差硬幣投擲法與帶權重上下法的閾值估計結果稍優於德爾曼上下法的模擬結果。

這些模擬結果具有現實意義：研究者在實驗設計階段選定回答信心的標準後，可根據這些結果選擇合適的步長與適測方法以達到較佳實驗結果。具體而言，當研究者設定回答信心標準較為嚴格時（即高截點的情境），若此標準偏離

回答信心的中點較遠，研究者可增大步長以減小初始值對閾值估計結果的影響。而根據三種適測方法的模擬結果，研究者最好優先選擇偏差硬幣投擲法與帶權重上下法，再考慮德爾曼上下法。

另一方面，本文亦考慮在非單調心理計量函數中應用適測方法估計閾值的情境，因此同樣分析在「同」/「異」作業裡各個適測方法的表現。與「是」/「否」作業具有類似結論，適測方法在「同」/「異」作業中的表現亦受目標概率與截點位置影響且具有相似的偏誤傾向；偏差硬幣投擲法與帶權重上下法的表現仍稍優於德爾曼上下法的模擬結果。差異在於在「同」/「異」作業中適測方法的模擬結果亦更易受初始值的影響；這表明研究者在設計「同」/「異」作業的實驗時更需注重初始值的合理選擇使其盡可能接近閾值。在適測方法的選擇上，研究者選擇偏差硬幣投擲法與帶權重上下法作為適測方法亦優於德爾曼上下法。

此外，在「同」/「異」作業中，本文亦比較兩種探討適測方法的方式：以Hsu與Doble (2015) 對心理計量函數的變換為基礎結合傳統適測方法，與García-Pérez (2014) 發展的新適測方法，並探討它們的差異性。其中，Hsu與Doble (2015) 結合傳統適測方法更適合閾值估計實驗，García-Pérez (2014) 的適測方法更適合估計完整心理計量函數。而García-Pérez (2014) 的適測方法與傳統適測方法亦有不少相似性，這些相似性或為將García-Pérez (2014) 的適測方法改用於閾值估計實驗提供可能性。

雖然García-Pérez (2014) 在其研究中指出採用有限多個回答信心選項會導致估計結果有較高偏誤；且本文在閾值估計實驗裡引入回答信心量尺後各種適測方法的估計結果亦存在偏誤，但本模擬實驗表明這種偏誤可通過增大試驗數或步長、調整初始值、改變適測方法等方式減小甚至完全消除（見目標概率為0.5時三種適測方法在「是」/「否」作業與「同」/「異」作業的模擬結果）。

更重要的是，在閾值估計實驗中引入回答信心量尺不僅能蒐集受試者每次回



答的信心程度，亦可在實驗設計階段給予研究者更多的自由。研究者可從實驗目的與需求出發，選擇合適的回答標準作為回答信心的截點，以決定受試者的回答信心應以何種標準判定（即，研究者設定高截點作為回答信心的標準可表明研究者希望以更為「嚴格」的標準認定受試者的肯定回答；若設定低截點的回答信心標準表明研究者希望以「寬鬆」的標準認定受試者的肯定回答。）。這種通過設定截點切割受試者的回答信心為肯定回答或否定回答的做法，它等同於原二選項回答模式，因此它可避免對傳統適用於二選項回答適測方法的修改。不僅如此，這種做法給予研究者更多自由決定如何篩選受試者回答，截點的可選擇性使研究者能自主選擇剔除或保留「低質量」（即受試者盲目猜測）回答。若此時實驗目的不變而實驗標準稍加變化，研究者只需修改截點位置而維持實驗設計流程不變，亦無需修改適測方法即可重新進行實驗。

此外，本文研究亦可用於現實生活：例如火警警報鈴聲的音量設定，研究者可將回答信心量尺的截點設定較高，此舉能確保受試者在一定概率下（即目標概率）能聽到該鈴聲；因此日常生活中，當他很確信聽到火警警報就可立即採取措施。反之，若研究者想了解某地施工噪聲應處於何種範圍內時，因噪聲對人有健康危害，此時研究者應設定較低截點以確保對多數人不會造成影響。類似地，藥物劑量的研究抑或過敏反應的測定，都適合採用較低截點做閾值估計實驗；而火警警報、地震警報等適合採用高截點做閾值估計實驗，確保多數人在危機時刻能聽到這些警報。

最後，本文僅採用電腦模擬方式評估各種適測方法在引入回答信心量尺的「是」/「否」作業與「同」/「異」作業中估計閾值的表現，未來仍需蒐集受試者實驗數據以探討引入這種量尺對閾值估計實驗的影響，及各種適測方法的表現差異。此外，本文選取三種適測方法均為無母數固定步長的適測方法，未來應考慮在添加這種回答信心量尺的閾值實驗裡採用不固定步長適測方法與有母數適測方法的表現。



第七章 參考文獻

- Chen, Y.-H., & Hsu, Y.-F. (2010). Convergence of non-parametric adaptive threshold estimation in the yes-no task. *Chinese Journal of Psychology, 52*, 95-112. doi: 10.6129/CJP.2010.5201.06
- Derman, C. (1957). Non-parametric up-and-down experimentation. *The Annals of Mathematical Statistics, 28*, 795-798. doi: 10.1214/aoms/1177706895
- Dixon, W. J., & Mood, A. M. (1948). A method for obtaining and analyzing sensitivity data. *Journal of the American Statistical Association, 43*, 109-126. doi: 10.1080/01621459.1948.10483254
- Durham, S. D., & Flournoy, N. (1995). Up-and-down designs I: Stationary treatment distributions. *Lecture Notes-Monograph Series, 25*, 139-157. doi: 10.1214/lnms/1215451483
- García-Pérez, M. A. (2014). Adaptive psychophysical methods for nonmonotonic psychometric functions. *Attention, Perception, & Psychophysics, 76*, 621-641. doi: 10.3758/s13414-013-0574-2
- García-Pérez, M. A., & Alcalá-Quintana, R. (2017). The indecision model of psychophysical performance in dual-presentation tasks: Parameter estimation and comparative analysis of response formats. *Frontiers in Psychology, 8*, 1142. doi: 10.3389/fpsyg.2017.01142
- Hsu, Y.-F., & Chin, C.-L. (2014). On the limitations of fixed-step-size adaptive meth-

- ods with response confidence. *British Journal of Mathematical and Statistical Psychology*, 67, 266-283. doi: 10.1111/bmsp.12018
- Hsu, Y.-F., & 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. doi: 10.1111/bmsp.12040
- Kaernbach, C. (1991). Simple adaptive testing with the weighted up-down method. *Perception & Psychophysics*, 49, 227-229. doi: 10.3758/BF03214307
- Kaernbach, C. (2001). Adaptive threshold estimation with unforced-choice tasks. *Perception & Psychophysics*, 63, 1377-1388. doi: 10.3758/BF03194549
- Klein, S. A. (2001). Measuring, estimating, and understanding the psychometric function: A commentary. *Perception & Psychophysics*, 63, 1421-1455. doi: 10.3758/BF03194552
- Leek, M. R. (2001). Adaptive procedures in psychophysical research. *Perception & Psychophysics*, 63, 1279-1292. doi: 10.3758/BF03194543
- Treutwein, B. (1995). Adaptive psychophysical procedures. *Vision Research*, 35, 2503-2522. doi: 10.1016/0042-6989(95)00016-X



第八章 附錄

附錄一 「是」 / 「否」 作業下均方根誤差與偏誤的最大值/最小值

表 1 「是」 / 「否」 作業下均方根誤差與偏誤的最大值/最小值

Method	Target	Cut-off	Min (RMSE)	Max (RMSE)	Min (bias)	Max (bias)
WUD	0.750	0.200	0.048	0.617	-0.043	0.617
	0.750	0.500	0.038	0.208	-0.177	0.192
	0.750	0.800	0.038	0.329	-0.316	0.049
BCD	0.750	0.200	0.059	0.637	-0.045	0.625
	0.750	0.500	0.043	0.219	-0.140	0.195
	0.750	0.800	0.049	0.294	-0.266	0.036
DUD	0.750	0.200	0.083	0.921	-0.041	0.902
	0.750	0.500	0.066	0.432	-0.162	0.403
	0.750	0.800	0.071	0.325	-0.303	0.081
WUD	0.500	0.200	0.043	0.229	-0.030	0.203
	0.500	0.500	0.040	0.115	-0.089	0.085
	0.500	0.800	0.044	0.229	-0.202	0.028
BCD	0.500	0.200	0.043	0.229	-0.032	0.204
	0.500	0.500	0.039	0.115	-0.086	0.085
	0.500	0.800	0.043	0.230	-0.205	0.026
DUD	0.500	0.200	0.043	0.228	-0.031	0.204
	0.500	0.500	0.039	0.116	-0.086	0.088
	0.500	0.800	0.044	0.227	-0.204	0.029

附錄二 「同」 / 「異」 作業下均方根誤差與偏誤的最大值/最小值



表 2 「同」 / 「異」 作業下均方根誤差與偏誤的最大值/最小值

Method	Target	Cut-off	Min (RMSE)	Max (RMSE)	Min (bias)	Max (bias)
WUD	0.750	0.200	0.049	0.933	-0.057	0.933
	0.750	0.500	0.038	0.525	-0.516	0.142
	0.750	0.800	0.027	0.417	-0.395	0.027
BCD	0.750	0.200	0.059	0.942	-0.051	0.934
	0.750	0.500	0.050	0.493	-0.469	0.144
	0.750	0.800	0.036	0.343	-0.297	0.020
DUD	0.750	0.200	0.084	1.256	-0.050	1.242
	0.750	0.500	0.071	0.510	-0.490	0.324
	0.750	0.800	0.053	0.397	-0.364	0.082
WUD	0.500	0.200	0.035	0.371	-0.017	0.334
	0.500	0.500	0.083	0.279	-0.261	0.250
	0.500	0.800	0.036	0.388	-0.350	0.009
BCD	0.500	0.200	0.035	0.365	-0.015	0.329
	0.500	0.500	0.081	0.279	-0.263	0.252
	0.500	0.800	0.035	0.380	-0.344	0.008
DUD	0.500	0.200	0.034	0.374	-0.016	0.335
	0.500	0.500	0.081	0.278	-0.258	0.250
	0.500	0.800	0.036	0.395	-0.360	0.006

附錄三 Hsu與Doble (2015) 方法在「同」 / 「異」 作業的應用

本模擬實驗以Hsu與Doble (2015) 的結果為基礎，推導引入回答信心量尺的「同」 / 「異」 作業所對應的心理計量函數。他們觀察到一個關係：若現有 m 回答信心選項，其信心程度從 m_1, \dots, m_k 以此降序減小，每個回答選項有其對應的概率 p_k 代表受試者選取該回答選項的概率。當比較刺激（本文以 x 表示比較刺激）

小於標準刺激（以 a 表示標準刺激）時，前 k 個回答信心的概率和與其相反條件（比較刺激大於標準刺激）的前 $m - k$ 個回答信心概率和是互補概率（即這兩個事件是互補事件），以表達式描述即（見等式(17)）：

$$P_{x-a<0}(\text{first } k \text{ confidence categories}) + P_{x-a>0}(\text{first } m-k \text{ confidence categories}) = 1 \quad (17)$$

此外，因回答選項的總數為 m ，因此這 m 個回答信心選項的概率和為1。以比較刺激大於標準刺激的情境為例，其等式為（見等式(18)）：

$$P_{x-a>0}(\text{last } k \text{ confidence categories}) + P_{x-a>0}(\text{first } m-k \text{ confidence categories}) = 1 \quad (18)$$

綜上，聯立兩個等式（等式(17)與等式(18)）：

$$P_{x-a<0}(\text{first } k \text{ confidence categories}) = P_{x-a>0}(\text{last } k \text{ confidence categories}) \quad (19)$$

由等式(19)可知，在比較刺激大於標準刺激的情境裡最後 k 個回答信心選項的概率和等同於在比較刺激小於標準刺激的情境裡前 k 個回答信心選項的概率和。因此，在此基礎上若將 m 個回答信心選項擴展至無限個回答信心選項，由於該回答信心選項是降序排列，它等同於本文的回答信心量尺的逆序形式。進一步地，本文假設在這樣的連續量尺下Hsu與Doble（2015）觀察結果依舊成立；但連續的回答信心量尺無法用選項個數切割成兩部分，在此用截點位置將回答信心量尺切割為兩部分。

因此，在比較刺激小於標準刺激的情境裡（以該情境為基準設定截點位置，記為cutoff，見等式(20)），等式(19)的前 k 個回答信心選項等價於回答信心量尺上超過截點的部分；該部分的回答信心概率和是其累積回答信心函數的右邊部分，即實驗裡被認定為肯定回答（ $Y_n = 1$ ）的情境。而在比較刺激大於標準刺激

的情境裡（見等式(21)），等式(19)的最後 k 個回答信心選項等價於回答信心量尺上未超過 $1 - cutoff$ 的部分；該部分的回答信心概率和是其累積回答信心函數的左邊部分。綜上，兩種情境的等式表達為（見等式(20)與等式(21)）：

$$P_{x-a<0}(response\ confidence > cutoff) = 1 - F_{x-a<0}(cutoff) \quad (20)$$

$$P_{x-a>0}(response\ confidence < 1 - cutoff) = F_{x-a>0}(1 - cutoff) \quad (21)$$

其中， $F_{x-a<0}$ 是在比較刺激小於標準刺激下，回答信心函數的累積分布函數；同理， $F_{x-a>0}$ 是在比較刺激大於標準刺激下，回答信心函數的累積分布函數。又因為本文假設等式(19)的觀察結果仍成立，則等式(22)亦成立：

$$P_{x-a<0}(response\ confidence > cutoff) = P_{x-a>0}(response\ confidence < 1 - cutoff) \quad (22)$$

聯立此三式（見等式(20), (21), (22)）可得（見等式(23)）：

$$1 - F_{x-a<0}(cutoff) = F_{x-a>0}(1 - cutoff) \quad (23)$$

因此，當比較刺激小於標準刺激時，其心理計量函數為：

$$P_{x-a<0}(response\ confidence > cutoff) = 1 - F_{x-a<0}(cutoff) = F_{x-a>0}(1 - cutoff) \quad (24)$$

而根據本模擬實驗假設（見等式(14)與等式(15)），當兩組不同刺激強度具有相同的絕對值時（即比較刺激A-標準刺激=標準刺激-比較刺激B），它們有相同的回答信心分布（即相同的回答信心期望值與標準差），因此它們存在如下關係（見等式(25)）：

$$F_{x-a<0}(1 - cutoff) = F_{x-a>0}(1 - cutoff) \quad (25)$$

聯立以上兩個等式（見等式(24)與等式(25)）可得：

$$P_{x-a < 0}(\text{response confidence} > \text{cutoff}) = F_{x-a < 0}(1 - \text{cutoff}) \quad (26)$$

因此，「同」/「異」作業的心理計量函數為：

$$\begin{aligned} P_{\text{cutoff}}(x - a) &= F_{x-a < 0}(1 - \text{cutoff}, u, \sigma) \quad \text{when } x < a, \\ &= 1 - F_{x-a > 0}(\text{cutoff}, u, \sigma) \quad \text{when } x \geq a \end{aligned} \quad (27)$$

其中，cutoff表示截點， $F_{x-a < 0}(1 - \text{cutoff}, u, \sigma)$ 是當比較刺激小於標準刺激時，均值為 u ，標準差為 σ 的累積回答信心函數在 $1 - \text{cutoff}$ 的累積函數值；其均值與標準差分別按等式(14)與等式(15)計算（其中 $i = x - a$ ）。

此外，本文在此需指出的是，儘管本模擬實驗設定的回答信心分布因其設定原因，當比較刺激與標準刺激的差值具有相同絕對值時，它們的回答信心分布一致，但該對稱性無法推論至心理計量函數，亦不可傳遞。另一方面，本文設定「對稱」回答信心分布是基於實驗設計的考慮：受試者在實驗時並不知曉比較刺激與標準刺激的位置，他僅需判斷左右刺激的差距；而比較刺激與標準刺激的位置是隨機分配的結果。因此在每次「同」/「異」作業的試驗裡，當受試者判斷左右刺激是否一致時，具有相同絕對差值的比較刺激與標準刺激（即兩種刺激強度）在受試者看來是一樣的情境。

