二、文獻探討

2.1 測驗理論

測驗理論(test theory)（或全稱叫「心理測驗理論」）是一種解釋測驗資料間實證關係(empirical relationships)的有系統的理論學說，它的發展，迄今已邁入不同的新紀元，測驗理論學者通常把它劃分成二大學派：一為古典測驗理論(classical test theory, CTT)——主要是以真實分數模式(true score model)為骨幹，亦即觀察分數等於真實分數與誤差分數之和；另一為現代測驗理論(modern test theory)——主要是以試題反應理論(item response theory, IRT)為架構。這兩派理論目前並行流通於測驗學界，但試題反應理論卻有後來居上，逐漸淩駕古典測驗理論之上，甚至進而取而代之之勢(余民寧，民80[12])。以下就古典測驗理論和現代測驗理論分述如下：

2.2 古典測驗理論內容

古典測驗理論又稱傳統測驗理論(conventional test theory)，或古典真分數理論(classical true score theory)。

古典真分數理論的假定主要在界定觀察分數、真分數及誤差分數三者間的關係。其主要者有三：

1. 觀察分數(X)是真分數(T)與誤差分數(E)的和。

   即　$X = T + E$

2. 誤差分數與真分數間無相關。

   $\rho_{TE} = 0$

3. 兩不同測驗之誤差分數間無相關存在。

   $\rho_{EE'} = 0$

古典測驗項目分析時，選擇題目的標準常依據：

1. 項目難度（item difficulty）
2. 鑑別力指數（index of discrimination）
3. 相關鑑別力指數（correlational index of discrimination）

項目難度就是答對某題目人數的百分比，也就是項目平均數(items mean)。難度指數常以P 表示。而鑑別力指數就是高分組的通過率P_H 減去低分組的通過率P_L，常以D 表示之。亦即：

$D = P_H - P_L$

其他相關鑑別力指數則有下列三種常被用來做為分析測驗的內部一致性（internal consistency）。

1. 點值兩列相關（point-biserial correlation）。
2. 兩列相關（biserial correlation）。
3. α 係數（coefficient alpha）。

三種方法都是利用受試者在某題的得分和受試者的總分間的相關程度來獲知某題

目與整個測驗的內部一致情形(王寶墉，民84[13])。

雖然古典測驗理論採用的計算公式簡單明瞭、淺顯易懂，但卻具有以下五項缺失(余民寧，民80[12])：

1. 古典測驗理論所採用的指標（例如：難度、鑑別度及信度等）都是樣本依賴指標。指標會因受試者樣本不同而產生不同之結果。
2. 每位受試者接受相同的測量標準誤指標。沒有考慮受試者的個別差異；對高、低能力兩極端的受試者而言，這種指標不合理且不準確。
3. 古典測驗理論對於非複本（nonparallel form）但功能相同的測驗所測得之成績，無法提供有意義的比較。
4. 古典測驗理論對於信度假設是建立在複本（parallel form）測量的概念假設上，然而這個假設往往不存在於實際測驗情境裡。複本測量的理論假設是行不通的。
5. 古典測驗理論假設得分相同的受試者，其能力必然相同，而事實上，得分相同的受試者，其能力估計值會因為反應組型之不同而可能不同。

2.3 現代測驗理論

現代測驗理論正是為了改進古典測驗理論的缺失而誕生，它具有下列幾項特點，這些特點正是古典測驗理論所無法具備的（Hambleton & Swaminathan，1985[14]; Hambleton et al.，1991[15]; Lord， 1980[16]）：

1. 現代測驗理論所採用的試題參數（item parameters）（如：難度、鑑別度及猜測度等參數），是一種不受樣本影響（sample-free）的指標。
2. 當代測驗理論能夠針對每位受試者，提供個別差異的測量誤差指標，而非單一相同的測量標準誤，因此能夠精確推估受試者的能力估計值。
3. 當代測驗理論可經由適用的同質性試題組成的分測驗，測量估計出受試者個人的能力，不受測驗的影響（test-free），並且對於不同受試者間的分數，亦可進行有意義的比較。
4. 當代測驗理論提出以試題訊息量（item information）及試卷訊息量（test information）的概念，來作為評定某個試題或整份試卷的測量準確性，倒有取代古典測驗理論的「信度」，作為評定試卷內部一致性指標之趨勢。
5. 當代測驗理論同時考慮受試者的反應組型與試題參數等特性，因此在估計個人能力時，除了能夠提供一個較精確的估計值外，對於原始得分相同的受試者，也往往給予不同的能力估計值。
6. 當代測驗理論所採用的的適合度考驗值（statistic of goodness-of-fit），可以提供考驗模式與資料間之適合度、受試者的反應是否為非尋常（unusual）等參考指標。

古典測驗理論雖然不夠嚴謹，但是理論淺顯易懂，便於在實際測驗情境（尤其小規模資料）實施；現代測驗理論雖然嚴謹，但理論艱深難懂，僅適用於大樣本測驗資料的分析（余民寧，民80[12]）。

2.4 試題反應理論(Item Response Theory, IRT)

近年來電腦測驗結合試題反應理論及適性測驗之優點，已經發展成電腦化適性測驗。這種施測方式乃是電腦依據受試者〝暫時〞(provisional)的能力估計值，自動選擇符合受試者暫時能力之題目來加以施測，且可隨時依據受試者在新呈現的題目答題的對與錯情形，更新受試者的能力估計，並據以選取下一題目，直到受試者能力之估計已相當地精確為止。此種施測方法，每位受試者受測之題目不同，施測時間及題數也可以不同（陳新豐，民87[11]）。而在CAT 中重要的能力估計的理論依據就是所謂的試題反應理論，也就是本節所要探討的主要內容。

2.4.1 基本概念

試題反應理論建立在兩個基本概念上：（1）受試者（examinee）在某一測驗試題上的表現情形,可由一組因素來加以預測或解釋，這組因素稱為潛在特質（latent traits）或能力（abilities）；（2）受試者的表現和這組潛在特質之間的關係,可透過一條連續性嚴格遞增（monotonically increasing）的函數來加以詮釋，此函數稱為試題特徵曲線（item characteristic curve，簡寫為ICC）。其實把不同能力的學生在某試題的得分值連接成線，便是某試題的試題特徵曲線。試題特徵曲線的涵義，即是某種潛在特質的程度與其在某一試題上正確反應的機率。潛在特質程度愈高則在某一試題上正確反應的機率愈大（余民寧，民81[17]）。

2.4.2 試題反應理論的基本假設（assumptions）

任何一條試題特徵曲線所代表的涵義是：答對某一試題的機率是由考生的能力與試題的特性所共同決定。因此，試題反應理論具有下列幾項基本假設，唯有在這些假設都成立的前提下，試題反應模式才能被用來分析所有的測驗資料。

1. 單向度

單向度（unidimensionality）是指測驗中的各個試題都測量到同一種共同的能力或潛在特質（余民寧，民81[17]）。如何才算是符合單向度假設呢？R. K. Hambleton, & H. Swaminathan，1985[15]認為只要測驗資料有一個主控因素或成份就可算符合，而這個主控因素便是特質或能力(王寶墉，民84[13])。最常用於考驗測驗單向性的方法就是因素分析法。而使用因素分析法的時機則有兩種作法(王寶墉，民84[13])：

1.1 模式選資料：先決定好要使用的IRT 模式再選合適的題目，此時即可利用因素分析獲得單向性資料。

1.2 資料選模式：有了資料再選適用的模式，因為資料不只由因素分析來考驗單向性，其他如速度因素、鑑別力、不變性等了解都可做為模式選取的考量。

現代測驗IRT 的假定屬於強假定（strong assumption）不同於古典測驗的弱假定

（weak assumption）測驗資料很難完全符合單向性的假定，但IRT 模式有其強韌性（robustness），也就是允許某種程度的違反假定。因此，假定的考驗不能做為判定現代測驗生死的依據，應將之視為題目的性質（王寶墉，民84[13]）。

2.局部獨立性

局部獨立性（local independence）就是對某受試能力而言，項目間無相關存在。局部表示只針對一個受試、一種測驗分數或某一受試能力θ 而言，而非對整體受試。獨立表示項目間無相關，也就是統計獨立（statistical independence）。也就是一個題目不能為另一個題目提供線索。局部獨立可以由因素分析考驗或使用卡方考驗每一配對題目間的顯著性（王寶墉，民84[13]）。

3. 非速度測驗

測驗的實施不是在速度限制下完成。選用試題反應模式時，這項基本假設必須被考慮。假設3隱含在單向度假設（余民寧，民81[17]）。

4. 知道—正確假設

如果考生知道某一試題的正確答案，他或她必然會答對該試題。假設4隱含在單向度假設（余民寧，民81[17]）。

2.4.3 試題反應理論的參數模式

試題反應理論以試題特徵函數表達受試者能力和測驗反應間之關係，因函數中所採用的參數個數不同，可被區分為不同的模式，常用的有單參數、雙參數及三參數等三種數學模式，各模式之試題特徵函數公式所示（古松民，民90[18]）：

- 單參數模式：$P_{ij}(\theta_j) = \dfrac{1}{1+\exp[-D(\theta-b_i)]}$ （1）
- 雙參數模式：$P_{ij}(\theta_j) = \dfrac{1}{1+\exp[-Da_i(\theta-b_i)]}$ （2）
- 三參數模式：$P_{ij}(\theta_j) = c_i + (1-c_i)\dfrac{1}{1+\exp[-Da_i(\theta-b_i)]}$ （3）

其中，

D=1.702；

i：題目編號；

j：受試者編號；

θ_j：第 j 位受試者之能力值；

a_i 、b_i 、c_i：第i 題的鑑別度參數、難度參數、猜測度參數；

$P_{ij}(\theta_j)$：能力θ_j 答對第i 題的機率函數，其圖形是一種S 形曲線，值介於 0 與 1 之間。

公式(1)(2)(3)代表常用的項目反應模式，每一種模式都依其採用的試題參數的數目多寡來命名，都僅是用於二分法計分(dichotomous)的反應資料（亦即，正確反應者登錄為1，錯誤反應者為0 的資料）。

2.4.3.1. 單參數模式（one parameter model)

根據公式(1)的定義，所繪出單參數試題特徵曲線如圖1 所示。試題難度參數b的位置正好座落在正確反應機率為0.5 時能力量尺（ability scale）上的點；換言之，當能力和試題難度相等時（即$\theta - b_i = 0$），考生答對某試題的機會只有百分之五十。愈困難的試題，其試題特徵曲線愈是座落在能力量尺的右方；反之，愈簡單的試題，其試題特徵曲線愈是座落在能力量尺的左方。試題難度參數有時又叫作位置參數（location parameter）。理論上，b值的大小介於±∞之間，但實際應用上，通常只取±2之間的範圍。單參數對數形模式（one parameter logistic model）相通於George Rasch(1960)的模式，因此又有Rasch模式之稱（余民寧，民81[19]）。

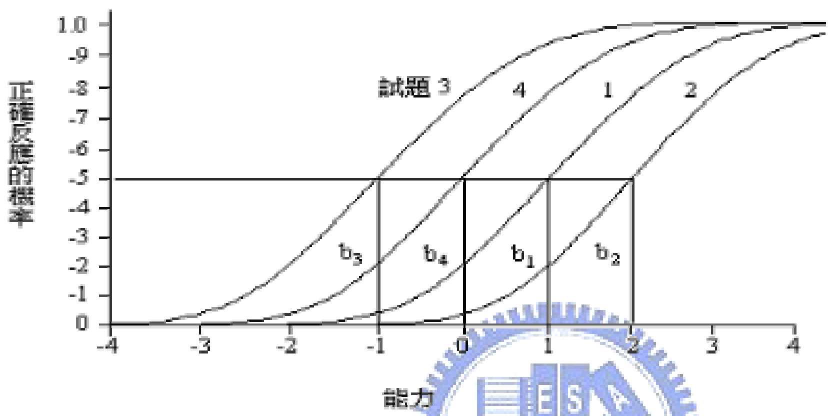

圖1 四條單參數模式的試題特徵曲線

資料來源：余民寧，民81[19]圖一

2.4.3.2. 雙參數模式(two parameters model)

與單參數模式相比，雙參數模式多了一個參數：試題鑑別度a(item discrimination)，根據公式(2)的定義，所繪出雙參數試題特徵曲線如圖2 所示。

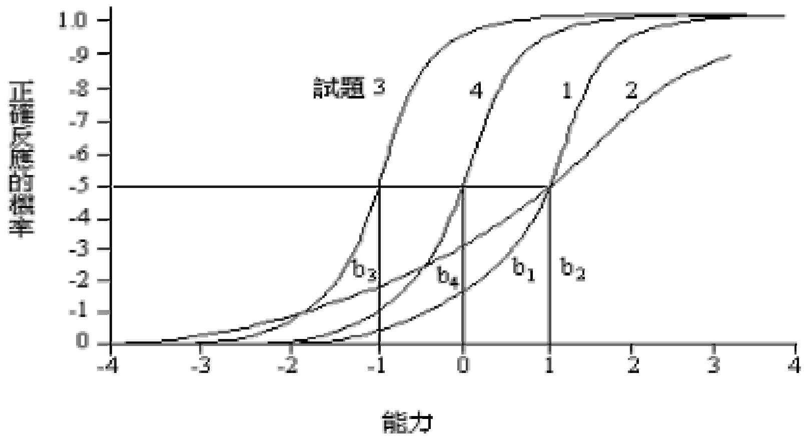

圖2 四條雙參數模式的試題特徵曲線

資料來源：余民寧，民81[19]圖二

試題鑑別度參數a 的值，剛好與在b 點的試題特徵曲線的斜率(slope)成某種比例。試題特徵曲線愈陡（steeper）的試題比稍平滑的試題，具有較大的鑑別度參數值；換句

話說，鑑別度愈大的試題，其區別出不同能力水準考生的功能愈好，亦即分辨的效果愈好。常用的a 值範圍介於0 與2 之間（余民寧，民81[19]）。

2.4.3.3. 三參數模式(three parameters model)

三參數模式則是加上了猜測度參數c，或稱機運參數（pseudo-chance parameter），這個參數提供試題特徵曲線一個大於零的下限，它代表著能力很低的考生答對某試題的機率。猜測參數的值比考生在完全隨機猜測下猜答的機率值稍小，亦即$C_i \leqq 1/A_i$，A_i 代表試題i 的選項數目，而C_i 為第i 題的猜測度參數。根據公式(3)的定義，所繪出三參數試題特徵曲線如圖3 所示。

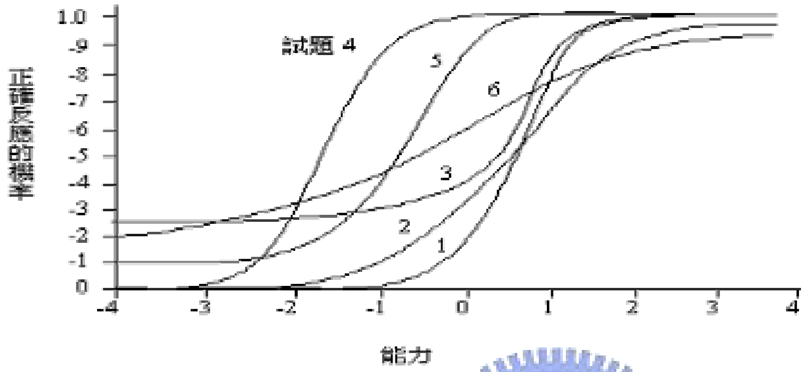

圖3 四條三參數模式的試題特徵曲線

資料來源：余民寧，民81[19]圖三

2.4.4 能力參數和試題參數的估計

試題參數與能力參數之估計又分為三種情況，分述如下：

1. 已知能力參數，估計試題參數：這是在受試者能力參數已知的情況下，求取試題參數的方法，通常採取最大概似估計法（maximum likelihood estimation），透過牛頓—拉弗森法 (Newton-Raphson method)來求出收斂值，收斂值即是欲求之試題參數。而其起始值可用二系列、點二系列、試題選項數目的倒數為a, b, c的初始值來進行牛頓迭代過程（Baker，1992[20]）。

2. 已知試題參數，估計能力參數：這是在題目參數已知下，求取受試者能力參數的方法，方法仍是最大概似估計法（maximum likelihood estimation），但若遇上無法收斂時（例：全對或全錯），亦即找不到最高點，則需用貝氏估計法（Bayesian Estimation），使能力變為常態先驗分配（prior distribution），便可克服最大概似函數無法收斂，找不到最高點的問題。

3. 試題參數與能力參數均未知，估計試題與能力參數：這是在受試者能力未知，而試題參數也未知時所採用的方法，有聯合最大概似估計法(Joint Maximum Likelihood Estimation, JMLE)、邊際最大概似估計法(Marginal Maximum Likelihood Estimation, MMLE)、邊際貝氏估計法(Marginal Bayesian Estimation, MBE)…等方法。先決定試題參數之初始值，再用其來估算能力參數，將所估算出來得能力參數用來估算試題參數，

將前後試題參數相比較，直到其差距在某一設定收斂標準下則結束，否則重覆進行這些步驟，直到小於收斂標準為止（Baker，1992[20]）。

2.4.5 試題訊息函數

試題反應理論提出一個能夠用來描述試題或測驗、挑選測驗試題、以及比較測驗的相對效能的實用方法，該方法即需要使用試題訊息函數（item information function)，作為建立、分析、與診斷測驗的主要參考依據。

試題訊息函數的定義如下：

$$I_i(\theta) = \frac{[P'_i(\theta)]^2}{P_i(\theta)Q_i(\theta)} \qquad i=1,2\ldots,n \tag{4}$$

其中$I_i(\theta)$代表試題i 在能力為θ 上所提供的訊息，$P_i'(\theta)$為在θ 點上的$P_i(\theta)$值的導數，而$P_i(\theta)$為能力θ 在試題 i 上的試題反應函數，$Q_i(\theta)=1- P_i(\theta)$。

以三參數對數模式為例，上面的公式可以化簡為（Lord，1980[16]；Birnbaum，1968[21]）：

$$I_i(\theta) = \frac{D^2 a_i^{\,2}(1-c_i)}{[c_i + \exp(Da_i(\theta - b_i))] \times [1+\exp(-Da_i(\theta-b_i))]^2} \tag{5}$$

從公式裡，我們很容易便可推知 a、b、和 c 參數在試題訊息函數中所扮演的角色：

1. 當 a 值較高時，訊息量也會較大。
2. 當 b 值愈接近θ 時，訊息量較大；反之，b 值愈遠離θ 時，訊息量則較小。
3. 當 c 值接近 0 時，訊息量則會增加。

Birnbaum指出，某個試題所提供的最大訊息量，剛好出現在能力參數為θ_{max} 的點上，θ_{max} 的值為：

$$\theta_{\max} = b_i + \frac{1}{Da_i}\ln\left[\frac{1}{2} + \frac{1}{2}\sqrt{1+8c_i}\right] \tag{6}$$

根據 Birnbaum 的推導，一份測驗的訊息函數（test information function）是指它在某一個θ 值上所提供的訊息量，該訊息量剛好是在θ 值上的試題訊息函數之總和，記作 I(θ)：

$$I(\theta) = \sum_{i=1}^{n} I_i(\theta) \tag{7}$$

在θ 值上的測驗訊息量與該能力的估計值的精確性成平方根反比，其符號記作：

$$SE(\hat{\theta}) = \frac{1}{\sqrt{I(\theta)}} \tag{8}$$

其中$SE(\hat{\theta})$稱作估計標準誤（standard error of estimation）。該項指標只要在能力參數的最大近似估計值求出後，便可計算得出。有了能力參數的最大近似估計值，並且也求出在θ 值上的測驗訊息之後，我們便可以估計信賴區間的方式來解釋能力估計值的涵義（余民寧，民 81[22]）。

2.5 類神經網路與試題反應理論

類神經網路（Artificial Neural Networks, ANNs）是類似動物神經結構的一個平行計算模式，即一種以電腦來模擬人類腦神經細胞網路的科學。類神經網路是由大量非線性的運算單元（即：神經元 neuron）和位於這些運算單元間的眾多連結（links）所組成。這些相互連結的處理單元通常是以平行且分散的方式運算且置放於整個網路結構之中，而整個 ANNs 的聚集形式就如同人類的大腦一般，可透過樣本或資料的訓練來展現出學習（learn）、回想（recall）、歸納推演（generalize）的能力。類神經網路（ANNs）在處理形式配套（pattern matching）、分類（classification）、函數近似（function approximation）、最佳化（optimization）及資料分類（data clustering）…等都有很好效果 （張斐章、張麗秋、黃浩倫，民 92[23]）。國內學者孫光天教授則將此項技術運用於題目反應理論的試題參數估計（孫光天，民 89[9]）。

2.5.1. 基本架構

從不同觀點來看，類神經網路有很多分類方式，其中一類是以網路連結架構來進行分類。以網路的連結架構來區分，類神經網路可分為前饋式（Feedforward）類神經網路與回饋式（Feedback）類神經網路兩類。神經元是類神經系統中最小的訊息處理單元，也是整個系統運作的基礎。神經單元模型，如圖 4 所示。它主要由三個部分組成：（張斐章、張麗秋、黃浩倫，民 92[23]）

1. 一組權重：權重W_i主要是模擬不同生物神經元間的連結強弱。
2. 輸入訊號疊加器：主要模擬生物神經元受多方刺激時細胞膜電位的總變化量。
3. 活化函數（轉換函數）：用來轉化輸入訊息疊加後的輸出值範圍，一般來說經過正規化（normalized）的輸出值通常在〔0, 1〕或是〔-1, 1〕。

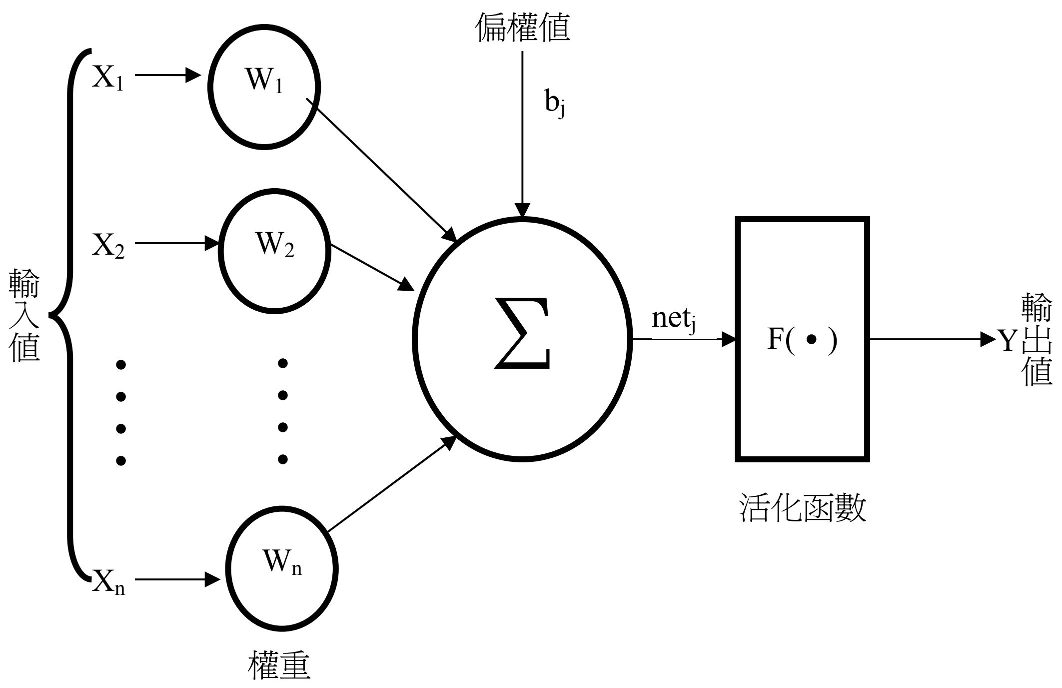

圖 4 類神經元模型

資料來源：張斐章、張麗秋、黃浩倫，類神經網路理論與實務，第 67 頁，民國 92 年。

就數學上而言，一個神經元 j 可以用下列二個方程式來描述：

$$net_j = \sum_{i=1}^{n} w_{ji} x_i + b_j$$ （作用函數 Activity Function） （9）

$Y_j = F(net_j)$ （轉換函數 Transfer Function） （10）

（張斐章、張麗秋、黃浩倫，民 92[23]）

前饋氏的類神經網路之架構的連結方式為單一方向的向前傳遞連結。最具代表性的前饋氏的類神經網路則是多層式認知類神經網路（multi-layer perception neural network）。圖 5 是多層式認知類神經網路的架構圖。（蔡志煌，民 89[10]）

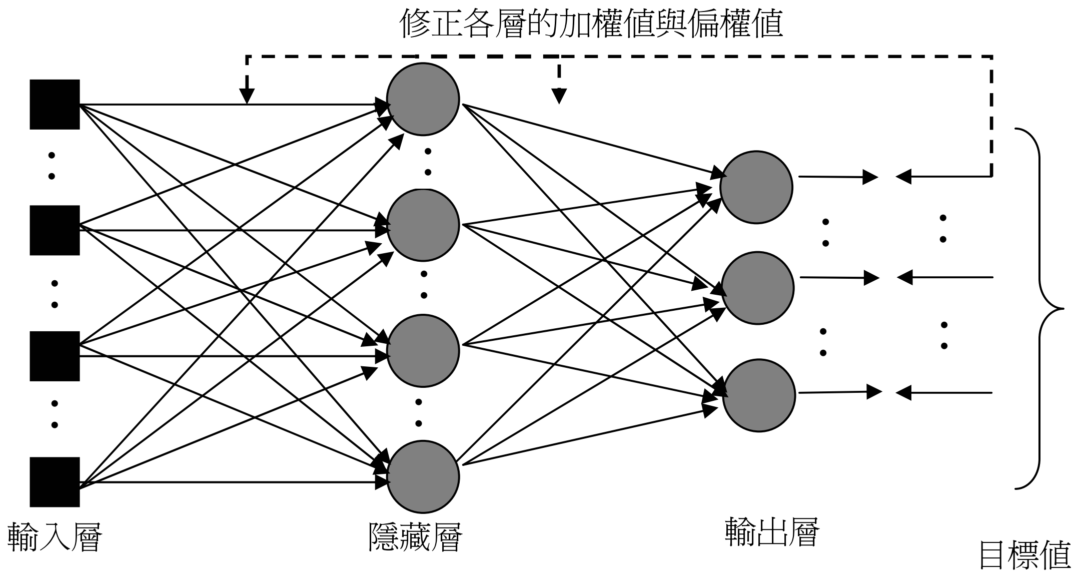

圖 5：多層式認知類神經網路的架構

資料來源：蔡志煌，「利用類神經網路於題目反應理論參數估計之研究」，國立台南師範學院，碩士論文，民國 89 年。

2.5.2 數學模式（轉換函數）--雙彎曲函數（sigmoid function）

雙彎曲函數（sigmoid function）$Y_j = \dfrac{1}{1+e^{-net_j}}$，　(11)

如圖 6 所示，是常用的轉換函數之一。而多層式認知類神經網路的架構可以用

$$f(x) = \frac{1}{1+e^{-(wx-t)}} \tag{12}$$

的數學模式表示。

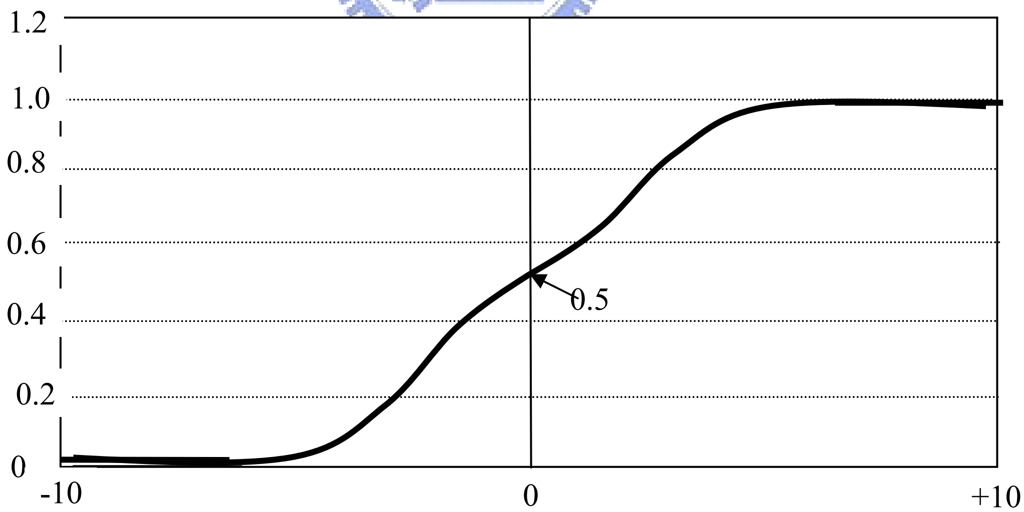

圖 6：雙彎曲函數，$F(v) = \dfrac{1}{1+e^{-\alpha v}}$，$\alpha = 1$

資料來源：張斐章、張麗秋、黃浩倫，類神經網路理論與實務的圖 2.16(c), alpha=1

2.5.3. 類神經網路的學習方式與學習演算法

類神經網路是以模擬大腦功能為主要目的，由於從經驗中學習是大腦特性之一，因此類神經網路中便利用各種不同學習演算法來模擬此一特性。通常我們以權重值的大小

來代表神經元間連結的強弱，如果從改變網路神經元間連結強弱的學習演算法來分類，大致可分成監督氏學習（supervised learning）與非監督式學習（unsupervised learning）二類。學習演算法就是一套權重調整演算法，藉由演算法逐步的調整神經元間連結的權重，使其達到最佳數值。有些情況下，權重值的調整主要是依據目標輸出值（desired output）與網路輸出值的差異大小而定。此一訓練過程因需要有目標輸出為標的，故稱為監督氏學習；反之，沒有目標輸出值的學習過程稱為非監督式學習。（張斐章、張麗秋、黃浩倫，民 92[23]）

倒傳遞類神經網路（Back Propagation Neural Network, BPN）的架構為多層認知器（multilayer perceptron, MLP），一般使用的學習演算法為誤差倒傳遞演算法（Error Back Propagation, EBP），簡稱為 BP（Back Propagation）演算法這樣的組合（MLP＋BP）稱之為倒傳遞類神經網路或 BPN（張斐章、張麗秋、黃浩倫，民 92[23]）。它是目前類神經網路學習模式中最具代表性，應用最普遍的模式。倒傳遞學習演算法屬於多層前饋式網路，以監督式學習方式，來處理輸出輸入間非線性映射關係。因為監督式學習旨在降低網路輸出單元目標輸出值與推論輸出值之差距，所以一般以下列能量函數（或稱誤差函數）表示學習的品質（葉怡成，民 92[24]）：

$$E = \frac{1}{2} \times \sum_{j} (T_j - A_j)^2 \tag{13}$$

其中　T_j=輸出層目標輸出值。

　　　A_j=輸出層推論輸出值。

因此網路學習過程變成使上式能量函數最小化的過程，通常以最陡坡降法（the gradient steepest descent method）來使能量函數最小化，即每當輸入一個訓練範例，網路即小幅調整加權值大小，調整幅度和誤差函數對該加權值的敏感程度成正比，即與誤差函數對加權值的偏微分值大小成正比（葉怡成，民 92[24]）。

$$\Delta W_{ij} = -\eta \bullet \frac{\partial E}{\partial W_{ij}} \tag{14}$$

，其中 η＝學習速率（learning rate），控制每次以最陡坡降法最小化誤差函數的步幅（葉怡成，民 92[24]）。

倒傳遞類神經網路運作原理主要可分為兩部份：

1. 學習過程：利用訓練範例的輸入向量與目標輸出向量，配合最陡坡降法將能量函數最小化，完成網路加權值的修正，此即學習過程的主要原理。學習過程通常以一次一個訓練範例的方式進行，直到學習完所有訓練範例，稱為一個學習循環（learning cycle），一個網路可以將訓練範例反覆學習數個學習循環，直到達到收斂（葉怡成，民 92[24]）。

2. 回想過程：為了檢驗學習的成果，通常在學習前的範例收集階段，將範例隨機分成二部份，一部份作為訓練範例，另一部份作為測試範例，在網路學習階段可每學習幾個學習循環，即將測試範例載入網路測試網路的誤差程度是否收斂？（葉怡成，民 92[24]）這是一種「分類」或「預測」的過程，可由訓練範例中找出與輸入最相近一個範例做為輸出。

2.5.4. 可調適輸出的類神經網路

2.5.4.1 類神經網路與二參數的數學模式

〝類神經網路模式與題目反應理論之雙參數的數學模式相仿，可以從以下數學式比較出來。

- 類神經網路的數學模式： $f(x)=\dfrac{1}{1+e^{-(wx-t)}}$
- 試題反應理論雙參數的數學模式： $P(\theta)=\dfrac{1}{1+e^{-1.7a(\theta-b)}}$

但因類神經網路的w和t值均是一個變動值，必須等到網路收斂時方可建立以下關係：

令x=θ，則 $f(\theta)=\dfrac{1}{1+e^{-(w\theta-t)}}$

∴ 1.7a=w => a=w/1.7 （15）

1.7ab=t => b=t/(1.7a)=t/w〞 （16）

（蔡志煌，民89[10]）

2.5.4.2 可調適輸出的類神經網路

〝由以上數學式可比較出，類神經網路與題目反應理論之二參數模式大致相同，但若要將類神經網路模型套在三參數模式中，則需要修改其數學式。首先在神經元上先增加一個輸入參數，此輸入參數稱之為下限參數(lower-floor parameter)，而此模式就稱之為可調適輸出的類神經網路〞，如圖7所示（引自蔡志煌，民89[10]）。

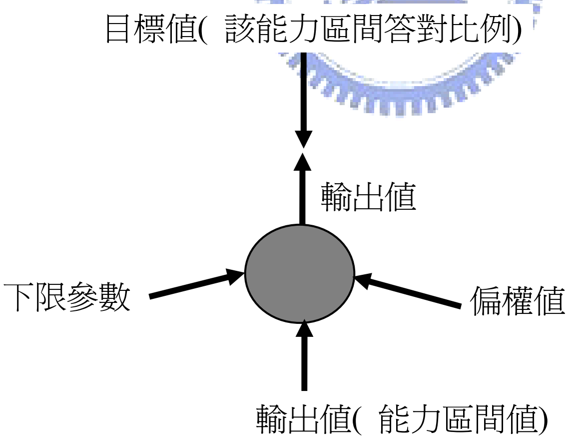

圖 7 可調適輸出類神經模式

資料來源：蔡志煌，「利用類神經網路於題目反應理論參數估計之研究」，國立台南師範學院，碩士論文，民國 89 年。

〝比較可調適輸出的類神經模式與三參數模式：

- 可調適輸出的類神經模式： $f(x)=c'+(1-c')\dfrac{1}{1+e^{-(wx-t)}}$ （17）
- 試題反應理論三參數模式： $P(\theta)=c+(1-c)\dfrac{1}{1+e^{-1.7a(\theta-b)}}$

由以上數學式可看出：

∴ $c=c'$ （18）

1.7a=w => a=w/1.7

$$1.7ab=t \quad => \quad b=t/(1.7a)=t/w$$

故a、b、c 三參數則可由可調適輸出的類神經模式來求得。待求得題目參數之後便可利用最大概似估計法來求受試者能力〞（引自蔡志煌，民89[10]）。

2.6　能力估計法(Ability Estimation Methods)

目前研究者最常採用的兩類受試能力估計方法為最大概似估計法及貝氏估計法。以下將對最大概似估計法(Maximum Likelihood Estimation, MLE)、貝氏順序估計法(Owen Sequential Bayesian Estimation, OWEN)、期望後驗估計法(Expected A Posterior , EAP)、和權值概似估計法（Weighted Likelihood Estimation, WLE）做簡要介紹：

2.6.1 最大概似估計法（Maximum Likelihood Estimation, MLE）

若能力為θ的受試者。回答第i個試題的機率密度函數記為$P_i(u_i|\theta)$，若作答正確，則$u_i = 1$；作答不正確，則$u_i = 0$。$P_i(u_i|\theta)$可以寫成

$$P_i(u_i|\theta) = P_i(u_i = 1|\theta)^{u_i} P_i(u_i = 0|\theta)^{1-u_i}$$

$$= P_i^{u_i}(1-P_i)^{1-u_i}$$

$$= P_i^{u_i} Q_i^{1-u_i} \tag{19}$$

，其中$P_i=P(u_i=1|\theta)$即為三參數對數模式中的$P_i(\theta)$，$Q_i(\theta)=1- P_i(\theta)$。

若施測的n個試題皆滿足局部獨立（local independence）的假設，則作答反應組型$u_1,u_2,\ldots u_n$之聯合機率為

$$P(u_1,u_2,\ldots,u_n|\theta) = P(u_1|\theta)P(u_2|\theta)\ldots P(u_n|\theta)$$

$$= \prod_{i=1}^{n} P(u_i|\theta)$$

$$= \prod_{i=1}^{n} P_i^{u_i} Q_i^{1-u_i} \tag{20}$$

若$u_1,u_2,\ldots u_n$ 分別為受試者（能力為θ）在n個試題的作答結果，則概似函數記為

$$L(u_1,\ldots,u_n|\theta)=\prod_{i=1}^{n} P_i^{u_i}(\theta) Q_i^{1-u_i}(\theta) \tag{21}$$

或 比較簡潔的表示式 $L(\mathbf{u}|\theta)=\prod_{i=1}^{n} P_i(u_i|\theta)$　(22)

因為概似函數值繪圖不便，通常將概似函數轉成自然對數的形式，即

$$L = \ln(u_1,u_2,\ldots,u_n|\theta) = \sum_{i=1}^{n}\left[u_i \ln P_i + (1-u_i)\ln Q_i\right] \tag{23}$$

接著L對θ作一次偏導數，

$$L_1 = \frac{\partial L}{\partial \theta} = \sum_{i=1}^{n} \frac{Da_i(P_i - c_i)(u_i - P_i)}{P_i(1-c_i)} = \sum_{i=1}^{n} K_i(u_i - P_i) \tag{24}$$

其中 $K_i = \dfrac{Da_i(P_i - c_i)}{P_i(1-c_i)}$ （25）

θ的最大概似估計值，$\hat{\theta}^M$ ，即為以下方程式的解：

$$\sum_{i=1}^{n} K_i(u_i - P_i) = 0 \tag{26}$$

由於上式並非線性程式，所以必須用牛頓迭代法(Newton-Raphson Iterative method)求解。首先給定初始值 $\hat{\theta}_0$ 為0，接著利用下式依序迭代得到能力估計值 $\hat{\theta}_1$ 、$\hat{\theta}_2$ 、…，

$$\hat{\theta}_{m+1} = \hat{\theta}_m - \frac{\left[\frac{\partial \ln L(u|\theta)}{\partial \theta}\right]_m}{\left[\frac{\partial^2 \ln L(u|\theta)}{\partial \theta^2}\right]_m} = \hat{\theta}_m - \frac{L_1}{L_2}\Big|_{\hat{\theta}_m} \tag{27}$$

，其中m=0, 1, 2, …. 為迭代次數；直到 $\left|\hat{\theta}_{m+1} - \hat{\theta}_m\right| \prec 0.001$ 才停止迭代。

在固定測驗長度（fixed test length）情況下，三參數對數型（3-PL）IRT模式的MLE能力估計偏差（bias）是

$$Bias[MLE(\theta)] = \frac{D}{I^2}\sum_{i=1}^{n} a_i I_i\left(\phi_i - \frac{1}{2}\right) \tag{28}$$

其中 $\phi_i \equiv \dfrac{P_i - c_i}{1 - c_i}$ （Lord，1983[25]）

又式(28)可以表示成 $Bias[MLE(\theta)] = \dfrac{1}{I^2}\sum_{i=1}^{n} Da_i I_i\left[\dfrac{1}{1+\exp[-Da_i(\theta - b_i)]} - \dfrac{1}{2}\right]$ （29）

從式(29)可看出MLE能力偏差接近零的直覺想法：任何能力的受試者都可從題庫中挑選出n個試題，每個試題難度接近該受試者能力估計值（targeted at an examinee's ability level）。在實際的CAT情境中，上述想法的涵意是：若題庫包含能夠估計幾乎所有能力值分佈且數目足夠大的試題量，則該CAT之MLE能力估計偏差應該是最小的（Yi et al.，2001[26]）。

2.6.2 貝氏順序估計法(OWEN Sequential Bayesian Estimation, OWEN)

Owen(1975)採用貝氏序列策略(Bayesian sequential approach)進行能力參數估計。在每一能力估計序列的步驟中，他假定能力參數的先驗分配為常態分配，而每一個先驗分配的平均數和變異數，即為前一次後驗分配的平均數和變異數。因為貝氏估計法將概似函數每一點的值乘上相對應常態分配的概率，即有峰值(peak value)出現，因此可以避免

能力估計值無法收斂的問題。但是將概似函數乘上常態分配之後，受試者的能力估計值會向先前分配的平均數迴歸，造成估計的偏差。貝氏法不會受制於全對或全錯反應題型，但卻會有迴歸效應—能力估計向其前驗分配的平均數迴歸（吳裕益等，民80[27]）。

2.6.3 期望後驗估計法（Expected A Posterior, EAP）

期望後驗估計法就是求後驗分配平均數。已知前驗分配 g(θ)，則後驗分配可以用(30)式表示。

$$p(\theta|u)=\frac{L(u|\theta)g(\theta)}{\int L(u|\theta)g(\theta)d\theta} \tag{30}$$

已知後驗分配 p(θ|u)，則後驗分配的平均值和變異數如下：

$$E(\theta\,|\,u)=\int_{-\infty}^{\infty}\theta p(\theta\,|\,u)d\theta \tag{31}$$

$$Var(\theta\,|\,u)=\int_{-\infty}^{\infty}\theta^{2}p(\theta\,|\,u)d\theta-\left(E(\theta\,|\,u)\right)^{2} \tag{32}$$

使用 Gauss-Hermite 能力點估計近似後驗分配的平均值和變異數，(31)、(32)式的兩個積分值可以近似如下

$$\hat{\theta}\equiv E(\theta|\mu)=\frac{\sum_{k=1}^{q}X_kL_j(X_k)W(X_k)}{\sum_{k=1}^{q}L_j(X_k)W(X_k)} \tag{33}$$

$$\hat{\sigma}^2(\hat{\theta})=Var(\theta|\mu)=\frac{\sum_{k=1}^{q}\left(X_k-\hat{\theta}\right)^2L_j(X_k)W(X_k)}{\sum_{k=1}^{q}L_j(X_k)W(X_k)} \tag{34}$$

(Bock & Mislevy，1982[28]；Wang & Vispoel，1998[29])

2.6.4 權值概似估計法（Weighted Likelihood Estimation, WLE）

Warm，1989[30]為了修正（correct）最大概似法所產生的能力偏差(bias)，提出一個 WLE 解決法：

$$L_1^{W}(\theta)=L_1(\theta)-\lambda(\theta)I(\theta)=0 \tag{35}$$

其中 $L_1=0$ 就是最大概似法的能力估計方程式，這個非線性方程式的解就是最大概似法能力估計值。(35)式中個別項目分別是：

$$L_1=\frac{\partial\ln L}{\partial\theta}=D\sum_{i=1}^{n}\frac{a_i\left[\mu_i-P_i(\theta)\right]\left[P_i(\theta)-c_i\right)]}{P_i(\theta)(1-c_i)}$$

，和$\lambda(\theta)=\frac{D}{I^2}\sum_{i=1}^{n}a_i I_i\left(\frac{P_i-c_i}{1-c_i}-\frac{1}{2}\right)$ （36）

，和$I(\theta)=\sum_{i=1}^{n}I_i(\theta)=\sum_{i=1}^{n}\frac{{P'_i}^2}{P_iQ_i}=\sum_{i=1}^{n}\frac{D^2{a_i}^2Q_i(P_i-c_i)^2}{(1-c_i)^2P_i}$ （37）

$\lambda(\theta)$是 MLE 的第 1 階能力估計偏差。根據 Warm 推論$\lambda(\theta)$可以藉由解(33)式加以消去，方程式的解稱為權值概似估計值（WLE, 以$\hat{\theta}$表示）(Cheng & Liou，2000[31] 、2003)[32]；Hambleton & Swaminathan，1985[15])。

由於式(33)並非線性方程式，所以同樣必須用牛頓迭代法求近似解。停止繼續迭代的判斷準則是〝$\left|\hat{\theta}^W{}_m-\hat{\theta}^W{}_{m+1}\right|\prec 0.001$〞或〝$\left|\hat{\theta}^W{}_{m+1}\right|\succ +5$〞或〝迭代次數達到 21 次〞三個條件中任一先發生（Warm，1989[30]）。

值得注意的是在 Warm（1989）主導的模擬研究中，他使用的是一個假想的（hypothetical）題庫（涵蓋大範圍 a、b 參數值），因為題庫的特性使得 MLE 能力估計的偏差值相當小，所以 WLE 的修正作用較難發揮（Yi et al.，2001[26]）。

2.7 測驗終止規則（Test-Termination Rule）對能力估計法的影響

Wang et al.，1999[33]指出在 CAT 環境中測驗終止規則（固定測驗長度 vs 變動測驗長度）可能會嚴重影響某些能力估計法的誤差指數（偏差、測量標準誤、均方根誤差）。例如：WLE 能力估計法原本是用來減少 MLE 能力估計法的估計偏差，但是 Wang et al.，1999[33]發現使用固定測驗長度終止規則時，WLE 能有效減少偏差；但使用固定標準誤（fixed-SE）終止規則，WLE 能力估計偏差大於 MLE（Yi et al.，2001[26]）。

Wang & Vispoel，1998[29]和 Wang et al.，1999[33]的研究中顯示在 CAT 環境中 MLE 能力估計法相較於貝氏能力估計法的估計偏差（bias）較小，但標準誤(standard error)較大。具有真實題庫的 CAT 採用固定測驗長度終止規則，其 MLE 的估計偏差會呈現兩端往外偏差（outward bias），也就是在 θ 量尺上較低能力端的能力值低估，較高能力端的能力值高估。另一種情況是採用 fixed-SE 終止規則時，則 MLE 的估計偏差會呈現兩端往內偏差（inward bias），也就是在 θ 量尺上較低能力端的能力值高估，較高能力端的能力值低估（Yi et al.，2001[26]）。