Journal of Propagations in Probability and Statistics
Vol. 2, No. 1, August 2001
Far Eastern Edition, pp. 1–6

統 計 薪 傳
第二卷　第一期
中華民國 90 年 8 月
遠東版　1–6 頁

最大概度估計之教學淺談

鄭惟厚
淡江大學

楔 子

在統計學的點估計理論中，最大概度估計是極為重要的方法之一。不過，對大專學生而言，此方法之原始觀念似乎稍嫌"抽象"，不易了解，也往往給授課教師們帶來困擾。有鑑於此，本期刊之編輯群特別邀請國內統計界以文筆流利著稱且教學經驗豐富的著名學者鄭惟厚教授跨刀為本刊執筆，藉由簡單的例子，說明最大概度估計背後的概念。鄭惟厚教授是暢銷書"統計，讓數字說話！"與"毛起來說 e"的譯者，此二書內容精彩有趣，經由鄭教授的巧譯之後，更是栩栩如生，美妙絕倫。我們不僅感謝鄭教授跨刀支持本期刊，更要向莘莘學子及統計界的教師與學者鄭重推薦上述二優良書籍，以提昇及促進統計之教育與薪傳。

(以上楔子由本期刊編輯委員撰寫)

□本文作者現任淡江大學數學系教授，與其兄鄭惟孝教授(Smiley W. Cheng，現任加拿大 University of Manitoba 統計系系主任)同為華裔統計界之著名學者。

教書多年以來有個很大的感觸，就是很多學生不喜歡思考，或不知如何思考。學東西不重概念，只重計算。連計算也只重結果(答案)，不重過程。只要算出來一個答案，好像就「交差」了。連答案明顯的不合理(比如算面積卻得到負數，或算機率結果卻大於 1)，有時都視而不見。而因爲只會方法不懂概念，就會出現一些有趣現象。你考一題難度頗高的題目，只要是「標準題型」的，一定有人會。要是以爲學生程度不錯而頗爲安慰，可就高興得太早了。下次考一題超級簡單，只要概念清楚，三分鐘就可以搞定的題目，結果是全倒。因爲不是標準題型，一般的方法著不上力。雖然只要懂定義就可以輕鬆做出來，無奈許多學生在學到方法後，就把定義打入冷宮不再青睞了！這樣子做其實是在捨本逐末，因爲

定義是最基本的概念。

基本概念重不重要呢?我想可以這樣說，如果基本觀念不清楚，只會一些方法的話，我們就是知其然而不知其所以然。雖然問題碰對了也可以解決，然而只要問題稍有一點變化的話，我們就只能瞪著它而束手無策了。舉個實際一點的例子好了。假如甲醫師基本觀念紮實，對他專攻的領域有充分的了解。乙醫師只學會基本動作，如此如此的症狀，就開如此如此的藥方。假如你感冒了，大概看哪位醫師都差不多。然而如果有疑難雜症，你願意把你自己交到哪位醫師手上呢?著重觀念還有附帶的好處呢。一旦你了解背後的意思，原本一堆符號、文字等等的組合，就會「有血有肉」起來，不再遙不可及。而且養成注重觀念的習慣後，自然而然會多思考，多問「什麼意思」和「爲什麼」。加強思考能力重不重要，我們來想想看。假如我們不會思考，遇到問題只會到記憶庫裡搜尋，看看是不是曾經處理過，那麼我們還有哪點比電腦強呢?在電腦越來越輕薄短小，使用越來越方便的今天，我們腦袋裡的記憶和搜尋功能已愈來愈不重要了。思考組織的能力，才是我們的最大本錢。世界上幾乎所有事物，有使用就有損耗。只有腦袋例外。愈用愈增長的東西，可千萬別省著用啊。

希望你已同意，了解觀念的重要。現在要來談我們的主題，最大概度估計了。當然我的重點會放在觀念上面。我們就是從定義概度函數(likelihood function)開始吧。這個概度函數，很多人稱它爲概似函數。我可不是喜歡標新立異。概度函數是國立編譯館編訂的「統計名詞」裡的翻譯，筆者還是編輯委員之一呢。我怕如果不用我們自己的「產品」，主任委員姚景星老師會來打我。好了言歸正傳。

定義 1 設 $X_1, X_2, ..., X_n$ 爲一來自某母體(population)的隨機樣本(random sample)，該母體之分佈，可用機率密度函數(p.d.f., probability density function) $f(x;\theta)$ 表示，θ 爲未知參數(parameter)，其可能範圍爲 Ω。則該樣本之概度函數定義爲

$$L(\theta) = L(\theta; x_1, x_2, ..., x_n) = f(x_1;\theta) f(x_2;\theta) \cdots \mathrm{f}(x_n;\theta) \text{，} \theta \in \Omega \text{。}$$

有兩點要補充說明。

1. 其實概度函數就是聯合密度函數(joint p.d.f.)，只不過主角換人做而已。聯合密度函數的主角是 $x_1, x_2, ..., x_n$，而概度函數的主角是 θ。
2. 定義中的 p.d.f.採廣義的解釋，也就是說不論分佈是連續的(continuous)還是離散的(discrete)，我們都稱爲 p.d.f.。所以我們的 p.d.f.也代表離散分佈的 p.m.f.(probability mass function)或 p.f.(probability function)。

接下來就要定義最大概度估計值(maximum likelihood estimate)了。

定義 2 固定一個樣本點(sample point) $\boldsymbol{x} = (x_1, x_2, ..., x_n)$，若 $\hat{\theta} = \hat{\theta}(\boldsymbol{x})$ 滿足 $L(\hat{\theta}) = \max\{L(\theta); \theta \in \Omega\}$，則稱 $\hat{\theta}(\boldsymbol{x})$ 爲 θ 之最大概度估計值。

又有幾點要補充。

1. 不同的統計書，定義的寫法或許不盡相同，但內容其實沒有實質上的差別。
2. 如果把 $\hat{\theta}(\boldsymbol{x})$ 中的小 x_i(是實數)用大 X_i(是隨機變數)來代替，所得到的 $\hat{\theta}(\boldsymbol{X})$，我們就稱它爲最大概似估計量(maximum likelihood estimator)。估計量是一個統計量，而估計值是一個數。沒取樣本之前我們只能找到估計量，取了樣本，把

樣本實際的值代進估計量，就有估計值了。

3. 如果用微積分的語言來說的話，$\hat{\theta}$ 其實就是 $L(\theta)$ 這個函數，在定義範圍 Ω 內，最大值發生的所在。

剛才說了半天，概念最重要。所以我們先利用一個簡單例子，來看看這個最大概度估計到底是什麼意思。

<u>**例 1**</u> 假設我請人鑄造了一個特別的銅板，故意造得不像一般銅板那樣均勻，所以就不能隨便假設出現正面的機率是 $1/2$ 了。我們把出現正面的機率叫做 p。現在把銅板給你，讓你擲 10 次，你會怎麼樣估計 p 呢？

讓我們先從直覺的角度看看這個問題。假設我們什麼統計方法都沒學過，或者是學過都忘光了(對於過了一個長長暑假的同學們，這個「境界」或許不難達到)，我們會怎麼來想這個問題呢？比如你擲的結果是反正反反正正反反反反，10 次當中，正面出現 3 次。應該用什麼來估計 p 呢？我們應該會說 3/10。爲什麼呢？因爲這是擲 10 次當中，正面出現的比例。才擲 10 次這麼少，樣本中正面出現的比例難道會接近真正的出現正面之機率 p 嗎？那當然不見得，可是若不用 3/10，難道有更理想的估計嗎？好像也想不出了。那我們就決定用 3/10 來估計 p 了，估且稱之爲直覺估計。好，現在暫時把這個「視窗」縮到最小，另開一個視窗來算一算，p 的最大概度估計會是什麼。

擲一個銅板，結果不是正面就是反面，所以我們面對的，是柏努利隨機變數(Bernoulli random variable)。現在把第 i 次擲銅板的結果，用 X_i 表示，令

$$X_i = \begin{cases} 1, & \text{若第 } i \text{ 次擲出正面}, \\ 0, & \text{若第 } i \text{ 次擲出反面}, \end{cases} \qquad i = 1, 2, \ldots, n.$$

在我們考慮的這個例子中，$n = 10$。而投擲的結果，可以用

$$(X_1, X_2, \cdots, X_{10}) = (0, 1, 0, 0, 1, 1, 0, 0, 0, 0)$$

表示。換個方式說，我們所觀測到的樣本點是 $\mathbf{x} = (x_1, x_2, \ldots, x_{10})$，而此處的 $x_2 = x_5 = x_6 = 1$，其他 x_i 均爲 0。

現在來找概度函數。對於柏努利分佈來說

$$f(x; p) = \mathrm{P}[X = x] = \begin{cases} p, & x = 1 \\ 1 - p, & x = 0 \end{cases} = p^x (1-p)^{1-x},\ 0 < p < 1.$$

所以概度函數就是

$$L(p) = f(x_1; p) f(x_2; p) \cdots f(x_n; p) = \prod_{i=1}^{n} p^{x_i} (1-p)^{1-x_i} = p^{\sum_{i=1}^{n} x_i} (1-p)^{n - \sum_{i=1}^{n} x_i}. \quad (1)$$

對我們的例子來說

$$L(p) = p^{\sum_{i=1}^{10} x_i} (1-p)^{10 - \sum_{i=1}^{10} x_i} = p^3 (1-p)^7,\ 0 < p < 1. \quad (2)$$

我們要找的最大概度估計值(MLE)，就是能夠使得(2)式中的 $L(p)$ 有最大值的 p

値。這是很普通的微積分問題。爲了方便計算，我們先取自然對數，因爲自然對數是遞增函數(increasing function)，它會把小的値對到小的，大的値對到大的，所以取了自然對數後，發生最大値的地方，和沒有取自然對數之前是一樣的。也就是取了自然對數後再找最大値發生的地方，所得到的 p 値，就是我們要的答案。現在取對數，得到 $\log L(p) = 3\log p + 7\log(1-p),\ 0 < p < 1$。微積分裏面說要找最大值，先得找出臨界點（critical points），也就是不能微分或微分等於 0 的點。這裏沒有不能微分的點，我們就找微分等於 0 的點。因此，

$$\frac{\partial \log L(p)}{\partial p} = \frac{3}{p} - \frac{7}{1-p} = \frac{3-10p}{p(1-p)} = 0 \quad \Rightarrow\ p = \frac{3}{10} = 0.3 \text{。}$$

這個 $p = 0.3$ 只是臨界點而已，嚴格上來說，應該檢驗一下，確認這是最大値發生的地方，我把這個檢驗的部份就留給讀者自己做。確認 $L(p)$ 在 $p = 0.3$ 有最大値之後，可知 p 的 MLE，就是 $\hat{p} = 0.3$ 了。把剛才縮小的視窗打開來看看，這個答案和我們剛才憑直覺做的，完全一樣，只是現在有根據。可是現在的做法，好像純粹是數學。而事實上 MLE 也是可以從很直觀的方向解釋的，只是我們常常把重點放到求解過程上，反倒忽略了 MLE 到底是什麼意思。現在我們再來換個方式看這個問題。

我們有一個銅板，不知道出現正面的機率 p 是多少。我們擲了 10 次，出現了 3 個正面（正面實際出現在什麼位置並不影響結果，只要總數同樣是 3，結果就一樣，這點從(1)式中就可以看出來）。現在我問一個問題，如果出現正面的機率 p 之實際值是 0.8，有沒有可能擲了 10 次出現 3 個正面？答案當然是可能的，但是你說機會不大，這我完全同意。如果 p 之實際值是 0.9 呢？仍然是可能的，只是機會更小了。那另一頭的情況呢？如果 p 之實際值是 0.1，可不可能擲 10 次出現 3 次正面呢？當然也可能啦。那現在就有件事得弄清楚了。既然我們現在手上有的資訊，就只有擲了 10 次得到 3 個正面；而不論真正的 p 是 0.1，0.8 還是 0.9，都有可能得到 3 個正面這樣的結果，那我們憑什麼決定用 $\hat{p} = 0.3$ 來估計 p，而不用 0.1，0.8 或 0.9 呢？因爲，你可能說，0.1，0.8 或 0.9 比較「不像」。不像的意思，是說如果 p 之實際值真的是比如說 0.8，那擲 10 次只出現 3 個正面的機會應該不大。好了，爲了具體一點，我們乾脆來算一些機率看看。當出現正面的機率爲 p 時，擲 10 次銅板恰好得到 3 次正面的機率是

$$\mathrm{P}\left[\sum_{i=1}^{10} X_i = 3\right] = \binom{10}{3} p^3 (1-p)^7. \tag{3}$$

現在把不同的 p 值代入上面的式子，或者查二項分布（binomial distribution）的機率表，然後把結果列出來(請見次頁之表 1)。從表一中的值我們知道，如果真正的 p 值是 0.8 的話，擲 10 次出現 3 個正面的機率是 0.0008。這個意思是說，如果我們連續地重覆擲銅板 10 次（此 10 次總稱爲一回合試驗）的試驗回合數很多很多的話，在這麼多試驗回合數中，出現恰好 3 個正面的機會，平均來說 10000 回合當中才只有 8 回。那難道這麼不容易發生的事情，我們現在才試驗一回合就發生了嗎？太不可能了對不對？所以我們當然不想用 0.8 來估計 p。怎麼樣的 p

才比較可能出現我們已觀測到的 3 個正面的結果呢？從表上來看，似乎就是 0.2，0.3 及 0.4，而其中又以 $p = 0.3$ 時，出現 3 個正面的機會最大。所以選 0.3 當做 p 的估計值，是不是很直觀的選擇呢？可是讀者諸君是否注意到，我們現在做決定的依據，其實就是最大概度估計的概念呢。我們剛才不是在找怎樣的 p 值，會使 $C(10,\ 3)\ p^3(1-p)^7$ 的值最大嗎？這和我們(2)式中的概度函數只差了 $C(10,\ 3)$ 這項，而這項是常數（不含 p），不影響結果。所以我們等於是在找使概度函數的值最大的 p 值，這不是 MLE 是什麼？當然為了顧到直觀，剛才我們稍稍"賴皮"，避開了微積分的部份。在比較 $C(10, 3)\ p^3(1-p)^7$ 之值的時候，我只算了 $p = 0.1$，0.2，…,0.9 這九種情況。可是誰敢保證從這裏面就可以找到最大值呢？怎麼知道代 $p = 0.31$ 到(3)式當中，就不會得到更大的值呢？如果你馬上代給我看，我就再問，那 $p = 0.305$ 呢？開始警覺了沒有，我可以沒完沒了的。要叫我閉嘴，就只能祭出微積分，才能大聲的說，沒錯的，$\hat{p} = 0.3$ 的確使(3)式有最大值，所以也使(2)式有最大值，所以就是最大概度估計！在結束這個例子之前，還有個小小的補充。細心的讀者可能有人注意到，表 1 當中當 $p = 0.9$ 時，得 3 個正面的機率居然是 0.0000，可是理論上不應該是 0 對不對？這是因為我們只取到小數第四位，若多取一位到小數第五位，你就會看到 0.00001 了。

表 1　不同 p 值之各個情況下擲銅板 10 次得 3 個正面的機率

p	擲 10 次得 3 個正面的機率	p	擲 10 次得 3 個正面的機率
0.1	0.0574	0.6	0.0425
0.2	0.2013	0.7	0.0090
0.3	0.2668	0.8	0.0008
0.4	0.2150	0.9	0.0000
0.5	0.1172		

為了強調概念的重要，我們再來看一個例子。

例 2 假設我面前有一個銅板，出現正面的機率不是 1/3 就是 2/3(記不清了)，我也記不得擲了兩次還是三次(年紀大了，沒辦法)，只知道總共擲出 2 個正面。所以我們面對的是一個二項隨機變數X，但是兩個參數都不知道。也就是說X的分佈為Binomial(n, p)，而n和p都未知，但我們知道$n = 2$ 或 3，$p = 1/3$ 或 2/3。現在請找出未知參數(n, p)之MLE。

好了你說，MLE 我會算。先寫出二項分布之 p.m.f.，即

$$\mathrm{P}[X = x] = \binom{n}{x} p^x (1-p)^{n-x}\ ,$$

然後概度函數就是

$$L(n, p) = \binom{n}{x} p^x (1-p)^{n-x} = \binom{n}{2} p^2 (1-p)^{n-2}\ ,\ 0 < p < 1\ ,$$

因為已知 $X = 2$。可是好像有點不大妙。兩個變數的函數找最大值，不是要分別對兩個變數做偏微分嗎？可是這裡的 $L(n, p)$ 怎麼對 n 偏微分呢？另外還有別的

傷腦筋的地方，就是 n 和 p 的範圍有很大的限制，各只有兩種可能。就算能用偏微分求出答案，之後這兩項限制要怎樣用上去呢？

如果我們只記得求 MLE 的方法，現在好像只能望題興歎了。可是我們如果記得定義，就不必這麼早放棄。最大概度估計的定義是什麼？是可以使得概度函數有最大值的參數值，不是嗎？現在概度函數已經有了，參數的可能值，總共也只有 $2\times 2=4$ 種可能的配對(n 只有兩種可能，p 也只有兩種可能)。我們光靠「土法煉鋼」，也就是把所有可能值一一代進概度函數，再比較看看哪一個值最大，不就可以得到結果了嗎？現在開始計算。

$$L(2,\tfrac{1}{3})=\binom{2}{2}\left(\tfrac{1}{3}\right)^2\left(\tfrac{2}{3}\right)^0=\frac{1}{9}\quad,\quad L(2,\tfrac{2}{3})=\binom{2}{2}\left(\tfrac{2}{3}\right)^2\left(\tfrac{1}{3}\right)^0=\frac{4}{9}\quad;$$

$$L(3,\tfrac{1}{3})=\binom{3}{2}\left(\tfrac{1}{3}\right)^2\left(\tfrac{2}{3}\right)^1=\frac{2}{9}\quad,\quad L(3,\tfrac{2}{3})=\binom{3}{2}\left(\tfrac{2}{3}\right)^2\left(\tfrac{1}{3}\right)^1=\frac{4}{9}\quad。$$

從計算的結果可以看出，在參數的所有可能值當中，使概度函數有最大值(即 $\frac{4}{9}$)的，共有兩組，所以 (n,p) 的 MLE 就是 $(\hat{n},\hat{p})=(2,\frac{2}{3})$ 或 $(3,\frac{2}{3})$。

這樣子的計算是不是超簡單的？而且 $\hat{p}=\frac{2}{3}$ 也不令人訝異，畢竟才擲 2 或 3 次，就得到兩個正面，所以正面機率應該較大才是。

希望上面的兩個例子，已幫助你對最大概度估計方法，有了更清楚的概念。例子都用離散分佈，是爲了直觀解釋的方便。一旦有了直觀解釋，是不是可以從一個新的角度來看最大概度問題，而不只是把它當作一個微積分問題了。

最後我們用一個較具挑戰性的例子收尾，但是我不想把答案先說出來，否則就沒樂趣了。(這裡說的樂趣可從兩方面來看，如果你很喜歡用腦筋解題，樂趣就是你的。如果你不大喜歡解題，我可以想像你抓耳搔腮，擠眉弄眼的模樣，那樂趣就是我的了，哈哈！) 題目是這樣的：在例 2 當中，如果 p 值仍然可能是 $\frac{1}{3}$ 或 $\frac{2}{3}$，但把 n 值「開放」，任何正整數都可以，擲的結果也仍然是兩個正面。那這題還解得出來嗎？我會給一些提示，但是我希望你先從直觀的方向想一想這個問題。比如，假如 p 是 $\frac{2}{3}$，那麼多擲幾次的話，正面共出現兩次的機率會比擲 2 或 3 次時的機率(即 $\frac{4}{9}$)高嗎？先想一想，猜猜答案，再去找答案，會有比較多的收穫。接下來討論如何解題。p 的可能值雖只有兩種，但 n 有無限多種可能，當然不能像例 2 一樣一一算出了。我給的提示如下(當然可能有別的做法)。我們可考慮先固定 p 值，比如 $p=\frac{1}{3}$，然後比較不同的 n 值之下，概度函數的變化情形，比如可以算出「後項比前項」，即

$$\frac{L(n+1,\frac{1}{3})}{L(n,\frac{1}{3})}=\frac{\binom{n+1}{2}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^{n+1-2}}{\binom{n}{2}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^{n-2}}\quad。$$

把上式化簡之後，討論一下它何時＞1，何時=1，何時＜1，就可能知道當 $p=\frac{1}{3}$ 時，$L(n,\frac{1}{3})$ 在 n 等於多少時有最大值了，對不對？那 $p=\frac{2}{3}$ 的時候呢？我再說的話，你就沒事做了，不能剝奪你的樂趣，祝你解題愉快(解答將刊登於本期刊下一期)。