檢定機率密度函數峰數的非參數方法

台灣大學數學系 鄭明燕

傳統上統計面對檢定一族群是否為均勻，即是否包含數個次族群的問題通常都是以參數方法解決。例如，使用數個單峰參數模型的混合分配以近似抽樣族群的分配並且檢定其配適度。這類參數方法缺點是它們受參數模型假設的正確與否影響極大。另外，即使在混合常態模型假設下，它們的配適度檢定的形式可能非常不好使用，而且可能必須倚賴高難度的數值計算方法。關於這方面的研究可參考 Hartigan [4] 及 Finch et al. [2]。

如果抽樣族群中各個次族群分別俱有單峰機率密度函數，並且區隔明顯，則這些次族群可經由非參數方法中之"bump hunting"方法，[例如 Good and Gaskins [3] 或檢定非參數機率密度函數估計是否為多峰的方法(例如 Silverman (1981) 的帶寬值檢定 (bandwidth test)] 加以確認。其中帶寬值檢定以核機率密度估計之帶寬值為檢定統計量。原則上，相對於前面提到的參數方法，這些非參數方法比較不受參數模型假設的影響。但是，那些經由非參數曲線估計建構的檢定可能受到其使用的曲線估計方法本身一些不合需要的特質所影響。例如，使用單一帶寬值的核機率密度估計量傾向由於樣本邊緣位置少數幾個觀察值而產生一些事實上不存在的小丘峰。這些問題降低帶寬值檢定對應用者的吸引力，並且降低檢定的檢定力 (power)。雖然這些問題可以由應用者直接介入調整來加以克服，但是如此便讓這檢定方法增加了一個主觀的因素，並且應用者並不一定俱備其中必要的統計技巧。

越質檢定統計量(excess mass test statistic)與 dip test statistic 分別由 Mueller and Sawitzki [6]及 Hartigan and Hartigan [4] 提出。在檢定一維的單峰分配時，前者恆為後者的兩倍，故此時以此兩者建立的檢定為等同。此統計量是經由直接比較樣本分配函數 (empirical distribution function) 與跟它最接近的具單峰機率之分配函數之間的差距得到的一個非參數檢定統計量。它的優點之一是它不須要倚賴非參數曲線估計。但是，它跟其他檢定峰數的非參數統計量有同樣的困難，就是很不容易以它建造一個俱有準確水準的檢定。其中難處在於如何對此統計量之值刻度 (亦即決定判斷點 (critical point) 的問題)。因為虛無假設下之所有單峰分配的集合構成一無限維度的函數空間。

Hartigan and Hartigan [4] 及 Mueller and Sawitzki [6] 提議的判斷點根據比較此統計量的樣本值與其均勻分配 (uniform distribution) 樣本的值觀察得到。這個辦法簡單易行但是得到的檢定過於保守。事實上我們可以證明當樣本數夠大時，給定任一非零的目標檢定水準 (nominal level)，由此原則得到的檢定的實際檢定水準會趨近於零。對照之下， Silverman [7] 以拔靴重抽樣 (bootstrap resampling) 原則刻度帶寬值檢定統計量建造的檢定，給定任一非零的目標檢定水準，具有非零的漸近實際檢定水準。這個結果可參考 Mammen et. al [5]。因此，如果在相同的目標檢定水準之下比較，當大樣本或樣本數適度時， Silverman 的檢定方法較傳統的越質檢定或帶寬值檢定擁有較大的檢定力。其實，Silverman 的檢定方法即使在大樣本之下也是相當保守的。這問題 Mammen et. al. [5] 已有探討，而 York [8] 亦給予數值印證，其檢定力並不理想。

近年來我與合作者的研究工作之一 (Cheng and Hall [1]) 便是提出一種刻度越質檢定量以檢定單峰分配的方法。我們的方法使用拔靴重抽樣來模仿在單峰虛無假設下越質檢定統計量之分配，由此達到刻度越質檢定之目的。它的檢定力可與 Silverman 的帶寬值檢定方法的檢定力匹敵，而且並不會如後者受到邊緣少數觀察值的影響。另外值得一提的是，我們建造此檢定的同時，也保證它具有準確的漸近檢定水準 (level)。以下簡述越質檢定統計量的定義、漸近分配以及如何刻度越質檢定以建造檢定方法。

假設 $c=\{X_1,\ldots,X_n\}$ 是一組從具有機率密度 f 的族群觀察到的樣本。這裡 n 是樣本數。令 F 為相對應於 f 的機率分配函數、$\hat{F}$ 是基於 c 得到的樣本分配函數。給定正整數 m 及正數 $l>0$，定義

$$E_{n,m} = \sup_{C_1,\ldots,C_m}\left[\sum_{j=1}^{m}\left\{\hat{F}\left(C_j\right) - l\left\|C_j\right\|\right\}\right],$$

這裡$\{C_1,\ldots,C_m\}$是不相交的間隔、$\hat{F}(C)$是C的$\hat{F}$測度、而$\|C\|$代表C的長度。定義

$$D_{n2} = E_{n,2}(l) - E_{n,1}(l) \geq 0。$$

則用以檢定f為單峰的虛無假設H_0，相對於f為多峰的對立假設H_1的越質檢定統計量定義為

$$\Delta_{n2} = \sup_{l>0}\{D_{n2}(l)\}。$$

如果Δ_{n2}的值太大，則應拒絕H_0而傾向接受f為非單峰的結論。刻度統計量Δ_{n2}的問題便是決定Δ_{n2}多大時應拒絕H_0。這須要Δ_{n2}在單峰虛無假設下的統計性質。

利用 Embedding 及一些分析核方法的技巧我們得到Δ_{n2}的大樣本漸近行為。其中最關鍵的是在單峰的虛無假設H_0下，Δ_{n2}的漸近分佈可表為兩項之乘積：一項為極複雜的隨機變量但與抽樣機率密度無關，另一項為一常數且完全決定於抽樣機率密度函數f及其微分在f的聚點(mode)的值。這裡一個函數的聚點代表此函數在此點有最大值。前面提到的常數為

$$c = \left\{f(x_0)^3 / \left|f''(x_0)\right|\right\}^{1/5},$$

其中x_0是f的唯一聚點。因此，我們希望找到一個單峰的‘刻度分配’F^0使得Δ_{n2}在F^0下與在F下，且虛無假設H_0為真時的統計性質相近。如此從F^0重抽樣Δ_{n2}的值便能決定檢定判斷點。

了解到$d = c^{-5} = \left|f''(x_0)\right| / f(x_0)^3$的所有可能值的集合為$(0,\infty)$。我們用以下三組參數分配含蓋此範圍：

(a) Beta(b,b)，$b>1$，分配：此時$d=g(b)$為b的單調遞增函數，且$d(1)=0$，$d(\infty)=2p$；

(b) 任一常態分配: 此時$d=2p$；

(c) rescaled student t-分配，具有機率密度

$$g_b(x) = \frac{1}{B(b-0.5,b)}\frac{1}{\left(1+x^2\right)^b}$$

$$,-\infty < x < \infty, \quad b > 1/2,$$

此時$d=g(b)$為b的單調遞減函數，且$d(1/2)=\infty$，$d(\infty)=2p$。

這三組參數分配的成員與d的所有可能值為一對一對應。下面介紹F^0的選取。

首先找到核密度估計值$\hat{f}$的聚點$\hat{x}_0 = \arg\max(\hat{f})$並且令

$$\hat{d} = \left|\hat{f}''(\hat{x}_0)\right| / \hat{f}(\hat{x}_0)^3$$

及$\hat{b} = g^{-1}(\hat{d})$。給定$c = \{X_1,\ldots,X_n\}$的條件分配，從機率密度$g_{\hat{b}}$的對應分配(令為$F^0$)，隨機抽取樣本$c^* = \{X_1^*,\ldots,X_n^*\}$。計算$\Delta_{n2}^*$，此處$\Delta_{n2}^*$為根據樣本$c^*$的$\Delta_{n2}$值。如此重複抽取$\Delta_{n2}^*$樣本值。則給定一目標檢定水準$a$，我們可利用 Monte Carlo 方法任意近似判斷點$\hat{z}_a$的值

$$P_{F^0}\left(\Delta_{n2}^* > \hat{z}_a \middle| c\right) = a,$$

並且於$\Delta_{n2} > \hat{z}_a$時拒絕虛無假設H_0。在一些極弱的條件下，這個檢定具有準確的漸進實際水準。從一個龐大的模擬實驗，我們驗證此結果並且得到下面的結論：

(a) 在虛無假設H_0下，我們的檢定比 Silverman 的方法具有更準確實際水準；

(b) 在對立假設H_1下，我們的檢定比 Silverman 的方法具有更佳的檢定力；

(c) 在虛無假設H_0或對立假設H_1下，即使f不像參數分配$g_{\hat{b}}$一樣為對稱函數，我們的檢定的實際水準或實際檢定力並不太受影響。

參考目錄

[1] M.-Y. Cheng and P. Hall, *J. R. Statist. Soc., B*. **60**, 579 (1998).

[2] S. J. Finch, N. R. Mendell and H. C. Thode, *J. Am. Statist. Ass.*, **84**, 1020 (1989).

[3] I. J. Good and R. A. Gaskins, *J. Am.. Statist. Ass.*, **75**, 42 (1980).

[4] J. A. Hartigan and P. M. Hartigan, *Ann. Statist.*, **13**, 70 (1985).

[5] E. Mammen, J. S. Marron and N. I. Fisher, *Probab. Theory Related Flds.*, **91**, 115 (1992).

[6] D. W. Mueller and G. Sawitzki, *J. Am. Statist. Ass.*, **86**, 738 (1991).

[7] B. W. Silverman, *J. R. Statist. Soc. B*, **43**, 97 (1981).
[8] M. G. York, Some problems in testing for modality, **Ph.D.** Thesis., Australian National University (1998).