

1. 前言

本文所要討論的是在抽樣調查中被視為隨機集合模型的樣本，並試著架構基於模糊統計邏輯的粗化數據 (coarse data) 之理論與特性。這個觀點跟 Nguyen and Wu (2006) 所提到的論點相當一致。也就是說有些抽樣調查中的數據可以視為隨機集合模型所得出的新數據，如何應用數學分析方法，配合軟計算技術以達到有效之資料處理與統計分析就是本研究之目的。

儘管抽樣調查中的隨機集合模型是人造的，在其他應用統計學領域中仍有一些自然隨機集合模型。一般時候，以修正過的數據或群組數據作為例子來說，自然隨機現象的觀察可能會有某些缺失，例如未能全面觀察或數據不精確或數據不完全。在這些個案中，很難直接由可見的母體 X 中分析隨機樣本 X_1, X_2, \dots, X_n 。近年來統計學家們竭盡所能的為要得到不可見樣本的相關資訊，創造了一種機制稱為粗略化機制。這對可見的樣本空間是一個必要的子集，並夾帶著不可見樣本 X_1, X_2, \dots, X_n 的相關訊息。正規地說，隨機粗略化機制就是一個隨機集合，而這種隨機集合的分佈，則是本文正在研究母體 X 未知分佈的函數，而並不是人造的。

隨機集合模型在統計學的某些領域中是自然而且實用的，雖然這個隨機集合的主題並不新穎，而且是完全被概括在機率理論的一般組織架構中，在機率理論和統計學的文章中通常很少提到主題。若排除隨機集合概念最早出現的時間點不談的話，它最早是由 Matheron (1975) 正式提出，在那之前，隨機集合的概念曾出現在幾何機率習題、幾何隨機抽樣和隨機模式的統計學分析中。關於隨機集合最近的幾篇文章請參閱 Goutsias、Mahler 和 Nguyen (1997), Nguyen (2002), Nguyen 和 Wang (2004) 和 Nguyen (2005) 提出的文章。

我們雖然知道一個透過隨機抽樣（也就是藉由某些人造隨機抽樣系統）獲得的樣本即為一個母體的字集。但是傳統上並沒有特別強調程序上的全域論點和隸屬度看法，因為抽樣實驗的結果有可能是某些集合的字集，而不是一個以數字表示出來的量。換句話說，必須將一個機率抽樣計畫視為一個隨機集合的分佈，也就是一個隨機元件取代某些集合的字集的數值。因此傳統隨機集合也就順理成章的成為抽樣調查中的模型。但是對於抽樣調查中的粗化數據，如何去

分析與解決，則是目前資料挖取 (data mining) 相當熱門之研究方向。

因此本研究將分析以下問題：

1. 建構將抽樣調查中最常見的粗化數據，做為一般統計學和資料分析模型的理論基礎。
2. 將一個機率抽樣計畫視為一個隨機集合模型的分佈，這有助於建構一個新的抽樣計畫。
3. 在探索抽樣調查的隨機集合和分佈的時候，機率測度論提供了很多種非精確數據給予統計學推測結論 (statistical inference) ，我們推廣傳統理論，以模糊集合及隸屬度為基礎，作為必要元素之運算據。