行政院國家科學委員會專題研究計畫成果報告

＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊
＊　　　　　　　　　　　　　　　　　　　　　　＊
＊混合機率分布於水文頻率分析適用性之研究（**II**）＊
＊　　　　　　　　　　　　　　　　　　　　　　＊
＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊

計畫編號：**NSC 89-2211-E-032-013**
執行期間：**88** 年 **8** 月 **1** 日至 **89** 年 **7** 月 **31** 日

計畫主持人：虞國興

執行單位：淡江大學水資源及環境工程研究所
中華民國八十九年七月三十一日

中文摘要

理論上混合機率分布之累積機率呈反曲或S型，由於水文資料常因其長度之不足，或樣本本身特性等因素，點繪於機率紙上時呈現反曲或S型的狀況，故易導致資料誤判為混合機率分布之結果。本研究將藉由混合機率分布參數特性之探討，提出一辨識資料所屬為單一機率分布抑或混合機率分布之檢驗方法。本研究以已知統計特性之合成資料，利用參數檢定方法分析當資料以混合機率分布擬合後，其分布參數間之特性，進而提出一分布檢驗方法，部份並與分布判斷準則 MSE 配合 F 檢定、分布群數判斷準則概似比率檢定（likelihood ratio test）作一比較。合成資料之產生兼以傳統上常用之單一機率分布（NOR、LN2、EV1 及 PT3）以及混合機率分布。研究中並針對台灣地區資料記錄年至少 30 年之年一日、二日及三日最大降雨量資料 250 站，進行分析，以探討本研究所提方法之適切性，並提供實際規劃之參考。研究結果顯示，以 MSE 配合 F 檢定作為分布之判斷準則時，單一機率分布之合成資料易被誤判為混合機率分布，誤判率達 85%以上。而本研究所提之分布檢驗方法，於單一機率分布之合成資料為小樣本時，精確度達 85%以上；於混合機率分布之合成資料為大樣本時，精確度可達 90%。而依據實測資料之分析結果可知，台灣地區年一日、二日、三日最大暴雨資料仍以單一機率分布較為適用。

英文摘要

The theoretical cumulative probability of mixture distributions also presents S or reverse curvatures on probability paper. For the real hydrological data, its curve shown on probability paper could be S or reverse curvatures due to the shortage of data or the various characteristics of data itself. Therefore, it is often to be misidentified as a mixture distribution. By probing into the property of the parameters of mixture distributions, this study proposes a parameter test method to identify whether the data is belonged to a single or a mixture probability distribution. Because the statistical properties of the synthetic data are known, this study examines the property of the parameters after the data is fitted by mixture distributions. Therefore, a distribution test method is proposed in this present study. The comparisons of the distribution test method based on MSE with F test and likelihood ratio test have been done in this study. Meanwhile, the synthetic data were generated both by single probability distribution (NOR, LN2, EV1 and PT3) and mixture probability distribution. In addition, the annual maximum 1-day, 2-day, and 3-day rainfall which have at least 30 years records from 250 stations in Taiwan area are used to inspect the aptness of the proposed test method that proposed in this study. The result shows that, a single probability distribution is more often to be misidentified as a mixture probability distribution when MSE with F test is chosen, the percentage of misidentification reaches 85%. On the other hand, the proposed test method provides 85% of accuracy to identify the single probability distribution of the synthetic data for a small sample size, while the accuracy could reach 90% by the mixture distributions for a large sample size. According to the results obtained from the real data, it is more appropriate to choose the single probability distribution for the annual maximum 1-day, 2-day and 3-day rainfall in Taiwan area.

謝　誌

本研究承蒙 行政院國家科學委員會之經費補助，得以順利完成，特此致謝。另對於淡江大學水資源及環境工程研究所研究助理曾世璋之協助，使本計畫順利進行，在此一並致謝。

目　錄

頁次

表目錄

頁次

圖目錄

頁次

第一章　緒論

1.1 研究動機與目的

各項興利除害之水利工程，自古以來即為民生必然之建設，但由於自然界中存在各種不確定因素，使得工程設計者必須在兼顧經濟及風險情況下，選擇較適當設計值為之，因此水資源工程中對於洪水頻率之精確估計，乃規劃設計與行政決策所亟需之重要依據。藉由統計分析，以過去之觀測記錄歸納其特性，推估未來可能發生之情況，即為決定水工結構物設計流量或雨量時所常用的水文頻率分析方法。在傳統水文頻率分析上，主要乃基於樣本係來自某一已知分布之母體之假設而進行研究，而決定何為其隱含之分布，則為頻率分析之首要工作。

水文資料常因資料長度不足，或樣本本身特性等因素，而在機率紙上呈現反曲或 S 型，此時機率分布之判定不易，故有混合機率分布的引進。混合機率分布乃由兩個或以上之母體以一權重因子所連結，在參數之不同組合（表組成母體間離散程度）與其個數差異的影響下，混合機率分布於資料擬合（data fitting）上往往較傳統單一機率分布能有更佳的表現。由於機率密度函數型式的較複雜，故其在分布參數與推估值的計算上亦較單一機率分布更為繁瑣費時。

虞及章（1998）發現，若於單一機率分布之合成資料分析時僅以均方誤差（MSE）為分布判斷準則，則最終決策以採用混合機率分布為多，誤判情形嚴重（達 90%以上），俟加以混合機率分布之群數判斷，雖可成功的將單一機率分布篩選出，大幅改善誤判情況（降至 10%左右），但卻僅侷限於少數單一機率分布（資料之原始分布為常態、對數常態分布時），此現象或因群數判斷準則概似比率檢定於使用上之限制所導致（Deb and Trivedi，1997；

Thode 等，1988；Atwood 等，1996）。據其研究之結果，單一機率分布之合成資料被誤判為混合機率分布之可能性偏高，而合成資料屬混合機率分布時之情況則未在研究範圍內。

有鑑於此，本研究之主要目的為找出一簡便且適宜之檢驗方法，用以辨識資料所屬為單一或混合機率分布，並一併探討合成資料為單一及混合機率分布時之判斷情形。

1.2 文獻回顧

某些集水區其洪水事件或有明顯不同的發生機制，此時可依發生之機制而將洪水事件分類，例如分為融雪流量及降雨流量兩類（Waylen and Woo，1982），或颱風流量及非颱風流量兩類，復進行頻率分析之。處理來自多於一個母體的資料，方法之一為採用混合機率分布進行頻率分析。混合機率分布乃由一權重因子連結數個分布元素所組成，分布元素之數目相依於母體數目。當混合機率分布以兩個母體連結時，亦稱為雙分布（Two – Distribution），又以雙常態分布、雙對數常態分布及雙指數分布較常被使用。其他混合型式如 Craigmile 及 Titterington（1997）以動差法與最大概似函數法推估參數之混合均勻分布；林及李（1991）於雙對數皮爾遜Ⅲ型分布之參數推估，比較最小平方法與最大概似法之差異。而，於不同之應用領域（如醫學上），其機率密度函數型式亦有所不同（Shen and Lin，1998；Wang 等，1998）。

Singh 等人（1972）最早將混合機率分布應用於水文統計分析，採用雙對數常態分布於洪水頻率分析，後並應用同樣概念於河川月流量資料之分析（1974）；Rossi 等人（1984）採用雙極端值分布於洪水頻率分析，模擬義大利流域之年洪峰流量；國內以林及涂（1988）最早使用雙對數常態分布，模擬濁水溪流域之桶頭站及集集站的月流量資料；林及李（1991）採用雙對數常態分布與雙對數皮爾遜Ⅲ型分布模擬月流量資料；虞及章（1998）探討混

合機率分布於水文頻率分析之適用性。

混合機率分布應用廣泛，其分布判定亦為一研究課題，相關研究有：Thode 等人（1988）對概似比率檢定（Likelihood ratio test）所使用之卡方分配其合適性進行探討；Atwood 等人（1996）針對概似比率檢定之檢定統計量與卡方分配之關係作一研究，提出應用於檢定時之卡方分配最適自由度應為 $2+\alpha n^{\beta}$，並在卡方分配自由度為 2 時提出一適用範圍；Chang（1976）探討具相同共變數矩陣之二維雙常態分布及其最適樣本大小之決定。

1.3 本研究架構

本研究所探討之主題可歸結如下：

(1) 藉由合成資料之分析，探討資料為單一機率分布及不同參數組合之混合機率分布時，以混合機率分布擬合所得之分布參數的特性。

(2) 藉由合成資料之分析，本研究提出一檢驗方法，以判別出資料所屬分布為單一機率分布或為混合機率分布。另，部份並與分布判斷準則 MSE 及分布群數判斷概似比率檢定作一比較。

(3) 針對不同樣本大小進行分析，探討資料長度所造成之影響。

(4) 藉由實測資料之分析，探討本研究所提方法之適切性。

本研究大綱為：第一章緒論；第二章理論基礎，分述頻率分析之理論與本研究之研究方法；第三章本研究所採用之資料，說明本研究使用之合成資料與實測資料；第四章結果與討論，討論合成資料與實測資料之分析結果；第五章結論。

第二章　理論基礎

本研究欲探討辨識資料所屬為單一機率分布或混合機率分布之檢驗方法。由於實測資料所屬分布為未知，故研究中以合成資料統計特性之已知，進行理論分析，提出一檢驗方法，續以實測資料分析實際應用之狀況。本章節為說明混合機率分布、四種常用單一機率分布、各分布參數推估法、反轉換值推衍、機率分布之判斷與檢定方法、合成資料之產生以及本研究之研究方法等。

2.1 混合機率分布及其參數推估

一機率密度可表為或可被擬合為一混合機率分布，即權重相加兩個以上不同參數之組成機率分布的總和，其具有多種組合，可以式(2.1)表示之：

$$f(x;\theta_g)=\sum_{i=1}^{g}p_i f_i(x;\beta_i) \tag{2.1}$$

其中，$\theta_g=\{p_1,\cdots,p_g,\beta_1,\cdots,\beta_g\}$，混合機率分布之參數集合

x：隨機變數

p_i：混合權重，$\sum_{i=1}^{g}p_i=1.$

β_i：組成密度之參數

$f_i(x;\beta_i)$：組成分布之機率密度函數

本研究所探討之混合機率分布為其中兩種狀況，即混合常態分布及混合對數常態分布。

2.1.1 混合常態分布

混合常態分布之機率密度函數如式（2.1.1）所示：

$$f(x) = pf_1(x) + (1-p)f_2(x) \tag{2.1.1}$$

其中 $f_1(x) = \frac{1}{\sqrt{2\pi}\sigma_1}\exp(-\frac{(x-\mu_1)^2}{2\sigma_1^2})$

$$f_2(x) = \frac{1}{\sqrt{2\pi}\sigma_2}\exp(-\frac{(x-\mu_2)^2}{2\sigma_2^2})$$

x：隨機變數

p：混合權重因子

混合常態分布之參數推估，本研究中採用 Dempster (1977) 極大化期望值演算法(expectation-maximization algorithm，簡稱 EM 方法)來計算最大概似估計值。假設已知 $x_i, i = 1,2,\ldots,n$，混合常態分布參數 p、μ_1、μ_2、$\sigma_1{}^2$ 及 $\sigma_2{}^2$ 之推估使用 Aitkin (1980)之迭代法, 計算流程如圖 1 所示（以 case1 為例）。其中，推估式依混合常態分布參數之不同組合而異，茲說明如下：

(1) 當 $\mu_1 \neq \mu_2, \sigma_1^2 = \sigma_2^2 = \sigma^2$ 時（以下簡稱 case1）

其 m（=4）個參數 p 、μ_1、μ_2 及 σ^2 之推估式如下：

$$\hat{p} = \sum_i \hat{a}(1|x_i) \Big/ n \tag{2.1.1a1}$$

$$\hat{\mu}_j = \sum_i x_i \hat{a}(j|x_i) \Big/ \sum_i \hat{a}(j|x_i) \quad j = 1,2 \tag{2.1.1a2}$$

$$\hat{\sigma}^2 = \sum_j \sum_i \{(x_i - \hat{\mu}_j)^2 \hat{a}(j|x_i)\} \Big/ n \tag{2.1.1a3}$$

其中，$\hat{a}(j|x_i)$：資料 x_i 屬於第 j 群機率分布之機率

n：樣本資料個數

(2) 當 $\mu_1 = \mu_2, \sigma_2^2 = k\sigma_1^2 \neq \sigma_1^2$ 時（以下簡稱 case2）

其 m（=4）個參數 p 、μ 、$\sigma_1{}^2$ 及 $\sigma_2{}^2$ 之推估式如下：

$$\hat{p} = \sum_i \hat{a}(1|x_i) \Big/ n \tag{2.1.1b1}$$

$$\hat{\mu}=\sum_{i}w_i x_i \Big/ \sum_{i} w_i \tag{2.1.1b2}$$

$$\hat{\sigma}_j^{\,2}=\sum_{j}\sum_{i}(x_i-\hat{\mu})^2 w_i \Big/ n \tag{2.1.1b3}$$

$$\hat{k}=\sum_{i} e_i^2 \hat{a}(2|x_i) \Big/ \sum_{i}\hat{a}(2|x_i) \tag{2.1.1b4}$$

其中， $w_i=\hat{a}(1|x_i)+\hat{a}(2|x_i)/\hat{k}$

$e_i=(x_i-\hat{\mu})^2/\hat{\sigma}_1$

(3) 當 $\mu_1\neq\mu_2,\sigma_1^2\neq\sigma_2^2$ 時（以下簡稱 case3）

其 m（= 5）個參數 p 、μ_1、μ_2、$\sigma_1{}^2$及$\sigma_2{}^2$之推估式如下：

$$\hat{p}=\sum_{i}\hat{a}(1|x_i) \Big/ n \tag{2.1.1c1}$$

$$\hat{\mu}_j=\sum_{i}x_i\hat{a}(j|x_i) \Big/ \sum_{i}\hat{a}(j|x_i) \quad j=1,2 \tag{2.1.1c2}$$

$$\hat{\sigma}_j^2=\sum_{i}\left\{(x_i-\hat{\mu}_j)^2\hat{a}(j|x_i)\right\} \Big/ \sum_{i}\hat{a}(j|x_i) \quad j=1,2 \tag{2.1.1c3}$$

2.1.2 混合對數常態分布

混合對數常態分布之機率密度函數如式（2.1.2）所示：

$$f(x)=pf_{y1}(x)+(1-p)f_{y2}(x) \tag{2.1.2}$$

其中 $f_{y1}(x)=\dfrac{1}{x\sqrt{2\pi}\sigma_{y1}}\exp(-\dfrac{\left(\ln x-\mu_{y1}\right)^2}{2\sigma_{y1}^2})$

$f_{y2}(x)=\dfrac{1}{x\sqrt{2\pi}\sigma_{y2}}\exp(-\dfrac{\left(\ln x-\mu_{y2}\right)^2}{2\sigma_{y2}^2})$

x：隨機變數

p：混合權重因子

參數 p、μ_{y1}、μ_{y2}、σ_{y1}^2及σ_{y2}^2之推估方法，與混合常態分布之三種狀況的推估式相同（$\mu_{y1} \neq \mu_{y2}$ ，$\sigma_{y1}^2 = \sigma_{y2}^2$，簡稱 case4；$\mu_{y1} = \mu_{y2}$ ，$\sigma_{y1}^2 \neq \sigma_{y2}^2$，簡稱 case5；$\mu_{y1} \neq \mu_{y2}, \sigma_{y1}^2 \neq \sigma_{y2}^2$，簡稱 case6），唯其中資料須先經過轉換，即令 $y_i = \ln x_i$，將 y_i 分別代入式(2.1.1a1)至式(2.1.1c3)等式中之 x_i，即可推估其不同狀況之參數。

2.2 四種常用單一機率分布及其參數推估

本研究共計採用常態分布、二參數對數常態分布、極端值 I 型分布及皮爾遜Ⅲ型分布，其中二參數對數常態分布之參數推估時資料不經轉換，皮爾遜Ⅲ型分布之參數推估時偏態係數須經修正，以下則分別說明之。

2.2.1 常態分布

常態分布之機率密度函數如式(2.2.1)所示：

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma_x} \exp(-\frac{\left(x - \mu_x\right)^2}{2\sigma_x^2}) \tag{2.2.1}$$

其中，μ_x及σ_x^2兩參數之推估值$\overline{x}$及$\hat{\sigma}_x^2$ 如式(2.2.1a)、(2.2.1b)所示：

$$\overline{x} = \sum_{i=1}^{n} x_i \Big/ n \tag{2.2.1a}$$

$$\hat{\sigma}_x^2 = \sum_{i=1}^{n} (x_i - \overline{x})^2 \Big/ n \tag{2.2.1b}$$

2.2.2 二參數對數常態分布

二參數對數常態分布之機率密度函數如式(2.2.2)所示：

$$f(x) = \frac{1}{x\sqrt{2\pi}\sigma_y} \exp(-\frac{\left(\ln x - \mu_y\right)^2}{2\sigma_y^2}) \tag{2.2.2}$$

其中，μ_y及σ_y^2兩參數之推估值$\hat{\mu}_y$及$\hat{\sigma}_y^2$ 採用資料不須轉換的方式求得，如式(2.2.2a)、(2.2.2b)所示：

$$\hat{\mu}_y = \ln(\overline{x}) - 0.5\ln\left[(\hat{\sigma}_x / \overline{x})^2 + 1\right] \tag{2.2.2a}$$

$$\hat{\sigma}_y^2 = \ln\left[(\hat{\sigma}_x / \overline{x})^2 + 1\right] \tag{2.2.2b}$$

2.2.3 極端值（最大）Ｉ型分布

極端值Ｉ（最大）型分布之機率密度函數如式(2.2.3)所示：

$$f(x) = \alpha \exp\left\{-\alpha(x-\beta) - \exp(-\alpha(x-\beta))\right\} \tag{2.2.3}$$

其中，α與β兩參數之推估值$\hat{\alpha}$與$\hat{\beta}$如式(2.2.3a)、(2.2.3b)所示：

$$\hat{\alpha} = 1.2825/\hat{\sigma}_x \tag{2.2.3a}$$

$$\hat{\beta} = \overline{x} - 0.45\hat{\sigma}_x \tag{2.2.3b}$$

2.2.4 皮爾遜Ⅲ型分布

皮爾遜Ⅲ型分布之機率密度函數如式(2.2.4)所示：

$$f(x) = \frac{1}{\alpha\Gamma(\beta)}\left(\frac{x-\theta}{\alpha}\right)^{\beta-1}\exp\left(-\frac{(x-\theta)}{\alpha}\right) \tag{2.2.4}$$

其中，β、α與θ三參數之推估值$\hat{\beta}$、$\hat{\alpha}$與$\hat{\theta}$如式(2.2.4a)、(2.2.4b)及(2.2.4c)所示：

$$\hat{\beta} = (2/\hat{C}_{Sx})^2 \tag{2.2.4a}$$

$$\hat{\alpha} = \hat{\sigma}_x \big/ \hat{\beta}^{0.5} \tag{2.2.4b}$$

$$\hat{\theta} = \overline{x} - \hat{\sigma}_x\hat{\beta}^{0.5} \tag{2.2.4c}$$

本研究採用偏態係數須經修正之方法，利用式(2.2.4d)修正得C'_{Sx}：

$$C'_{Sx} = C_{Sx}\frac{\left[n(n-1)\right]^{0.5}}{(n-2)}(1+8.5/n) \tag{2.2.4d}$$

復將C'_{Sx}之修正值代入式(2.3.4a)、(2.3.4b)及(2.3.4c)，即可求出α、β與θ之推估值。

2.3 反轉換值推衍

2.3.1 點繪法

各模擬分布所對應之推估值$\hat{x}_i$，係利用點繪法求出與各資料點x_i對應之累積機率，再進行反求而得。於點繪法之選擇，本研究採用虞及黃（1992）利用隨機變數順序統計量之中位數所推求之無關機率分布點繪法公式（Distribution Free Plotting Position）：

$$P_i = \frac{i-0.326}{n+0.348} \tag{2.3.1}$$

其中，i：水文資料之大小次序（i=1 時為最大）

n：水文資料個數

以隨機變數順序統計量之均數與眾數所推求之點繪法公式，隨機率分布的不同而有所不同，然而以隨機變數順序統計量之中位數所推求之點繪法公式，則與機率分布無關，可適用於所有機率分布。

2.3.2 推衍方法

混合常態分布之反函數值較為複雜，根據下列步驟說明如下：

1. 分別依照不同狀況求出其參數$\hat{\mu}_1$、$\hat{\mu}_2$、$\hat{\sigma}_1^2$、$\hat{\sigma}_2^2$及$\hat{P}$以及資料本身之$\bar{x}$、$\hat{\sigma}$。

2. 以資料之$B=\bar{x}+2.5\hat{\sigma}$為起始值，分別計算$Z_j=\dfrac{B-\hat{\mu}_j}{\hat{\sigma}_j}$，j=1,2。

3. 將Z_1、Z_2經反轉換求得對應之標準常態分布之累積機率a_1及a_2。

4. 計算於B點之雙常態分布的累積機率值$P'=\hat{P}a_1+(1-P)a_2$。

5. 若$|P'-(1-P_i)|<\varepsilon$則$x_i=B$（P_i表示所欲推求之反函數值所對應之超越機率，根據廣氏點繪法求得）。

6. 若$|P'-(1-P_i)|>\varepsilon$則$B=B-\varepsilon 1$，重回第2步驟直到$|P'-(1-P_i)|<\varepsilon$（本研究ε值採用0.0005，$\varepsilon 1$值視情況改變）。

其中，起始值$B=\bar{x}+2.5\hat{\sigma}$視兩組成分布之離散程度做調整（離散程度大則B值亦須較大），並將疊代間距$\varepsilon 1$在兩分布重合區間加大，則有助於縮短程式運行時間。混合對數常態分布反轉換值之求法與對數常態分布類似，先對所有資料取對數，代入上述之推求流程求得反轉換值，再計算其反對數轉換即為所求。

2.4 機率分布之判斷與檢定方法

2.4.1 分布判斷準則與F檢定

分布判斷準則採用MSE，其定義如下：

$$MSE=\frac{1}{n}\sum(\hat{x}_i-x_i)^2$$

其中，x_i：第i個隨機變數值

$\hat{x}_i$：第i個擬合機率分布之推估值

判斷MSE之差異是否達顯著，本研究採用F分配檢定，源於迴歸分析中對線性統計模式的檢定，方法如下：

full model： $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ （在此表示混合機率分布）

reduce model： $Y_i' = \beta_0 + \varepsilon_i$ （在此表示單一機率分布）

$H_0 : \beta_1 = 0$ （在此表示資料為單一機率分布）

$H_1 : \beta_1 \neq 0$ （在此表示資料為混合機率分布）

$$F^* = \frac{SSE(R) - SSE(F)}{df_R - df_F} \bigg/ \frac{SSE(F)}{df_R}$$

其中，$SSE = \sum (\hat{x}_i - x_i)^2 = MSE \times n$，誤差平方和（error sum of square）

df_F：full model（混合機率分布）之自由度

df_R：reduce model（單一機率分布）之自由度

於 SSE(F)≦SSE(R)時進行檢定，顯著水準（Level of Significance）α，分子自由度$df_R - df_F$，分母自由度df_R，決策法則（拒絕H_0之條件）為：

$$F^* > F_{1-\alpha; df_R - df_F, df_F}$$

2.4.2 混合機率分布群數判斷

以概似比率檢定（Likelihood ratio test）為混合機率分布之群數判斷，檢定方法如下：

$H_0 : g = g_1 = 1$ 表示資料為單一機率分布

$H_1 : g = g_2 = 2$ 表示資料為混合機率分布

$$-2c \log \lambda \sim \chi_v^2$$

其中，$-2\log\lambda = -2\left[L(g_1) - L(g_2)\right]$

$L(g_k)=\sum_{i=1}^{n}\ln\left[\sum_{j=1}^{g_k}\hat{p}_j f(x_i,\hat{a}_j)\right]$，$g_k$群時之概似函數值，k=1,2

χ_v^2：卡方（chi-square）分布，自由度 $v=2(m-1-d)$

$\hat{p}_j$：第 j 群推估之權重因子

$\hat{a}_j$：第 j 群推估之參數

c：$(n-2-\frac{1}{2}g_2)/n$

n：資料個數

d：單一機率分布之參數個數

m：混合機率分布之參數個數

2.4.3 雙母體平均數檢定

兩不同常態母體平均數μ_x、μ_y之檢定，本研究採兩母體變異數σ_x^2、σ_y^2為未知但相等時之檢定方法，則共同變異數σ^2可以其不偏估計量S_p^2代之，其中n_x、n_y分別為抽樣自兩母體之資料個數，$\hat{S}_x^2$、$\hat{S}_y^2$分別為σ_x^2、σ_y^2之推估值：

$$S_p^2=\left[(n_x-1)\hat{S}_x^2+(n_y-1)\hat{S}_y^2\right]/(n_x+n_y-2)$$

假設的建立與檢定統計量 T 如下所示，其中$\overline{X}$、$\overline{Y}$分別為μ_x、μ_y之推估值：

$$H_0:\mu_x-\mu_y=\delta_0$$

$$H_1:\mu_x-\mu_y\neq\delta_0$$

$$T=\frac{\overline{X}-\overline{Y}-(\mu_x-\mu_y)}{S_p\sqrt{\frac{1}{n_x}+\frac{1}{n_y}}}$$

透過自由度$\nu=(n_x+n_y-2)$，顯著水準α的 t 分配進行$\mu_x-\mu_y$之雙尾檢定，拒絕H_0之條件為：

$$t=\frac{\overline{X}-\overline{Y}-(\mu_x-\mu_y)}{S_p\sqrt{\frac{1}{n_x}+\frac{1}{n_y}}}=\frac{\overline{X}-\overline{Y}-\delta_0}{S_p\sqrt{\frac{1}{n_x}+\frac{1}{n_y}}}\geq t_{1-\alpha/2,\nu}\quad 或\quad \leq t_{\alpha/2,\nu}$$

2.4.4 雙母體變異數檢定

兩不同常態母體變異數$\sigma_x{}^2$、$\sigma_y{}^2$之檢定，透過分子自由度$\nu_x=(n_x-1)$，分母自由度$\nu_y=(n_y-1)$，顯著水準α的 F 分配進行之：

$$H_0:\sigma_x{}^2=\sigma_y{}^2$$

$$H_1:\sigma_x{}^2\neq\sigma_y{}^2$$

$$F=\hat{S}_x{}^2/\hat{S}_y{}^2$$

拒絕H_0之條件為：

$$f=\hat{S}_x{}^2/\hat{S}_y{}^2\geq f_{1-\alpha/2,\nu_x,\nu_y}\quad 或\quad \leq 1/f_{1-\alpha/2,\nu_x,\nu_y}$$

2.5 合成資料之產生

2.5.1 混合機率分布

以混合常態分布產生代表混合機率分布之合成資料，若總資料個數為 n，則方法步驟如下（case3 為例，分布參數 p、μ_1、μ_2、σ_1^2 及 σ_2^2）：

1. $n_1 = (p \times n)$，為第一組成分布之資料個數。
2. $n_2 = (1-p) \times n$，為第二組成分布之資料個數。
3. 產生常態分布具平均值 μ_1 及變異數 σ_1^2 之隨機變數 $x_1, x_2 \cdots, x_{n_1}$。
4. 產生常態分布具平均值 μ_2 及變異數 σ_2^2 之隨機變數 $x_{n_1+1}, x_{n_1+2}, \cdots$
   $\cdots, x_{n_1+n_2}$。
5. $n = n_1 + n_2$，$x_i, i = 1,2,\cdots n$ 即為所求。

為了解以混合機率分布擬合資料後所推得參數之變化，故加以考慮合成資料屬混合機率分布時之各種狀況。故，於組成分布所佔比例方面，變化權重因子 p=0.1（0.1）0.5。在 case1、case3 之組成分布間離散程度方面，以式(2.5.1)作為標準，分為三種狀況（考慮大於、等於及小於）做探討：

$$|\mu_2 - \mu_1| = 2(\sigma_1 + \sigma_2) \tag{2.5.1}$$

另，變化資料長度探討其影響，樣本大小取 n = 30（10）100（50）150。故，計有 3（case）x5（p 值）x3（離散程度）=45 種情況，每一情況包含 9 種樣本大小，每一樣本大小均產生 100 組。

2.5.2 四種常用單一機率分布

共計產生常態分布（NOR）、對數常態分布（LN2）、極端值（最大）Ⅰ型分布（EV1）及皮爾遜Ⅲ型分布（PT3）等四種分布，利用 IMSL Library Subroutine DRNNOR、DRNLNL、DRNUM、DRNGAM 等，關於各分布之參數推估方法見 Kite（1977）。以樣本個數 n = 30（10）100（50）150 產生每一分布之合成資料各 100 組。於探討擬合混合機率分布之參數特性時，並將資料長度加大（n=30（10）300），產生樣本組數增多至 1000 組。

2.6 研究方法

本研究利用單一及混合機率分布所產生之合成資料來模擬實測資料擬合之狀況，並分析判斷準則 MSE 配合 F 檢定、群數判斷概似比率檢定、雙母體平均值檢定及雙母體變異數檢定之判斷結果，以提出一適當之檢驗方法，用於混合機率分布與單一機率分布之判定。此外，並針對台灣地區之年一日、二日及三日最大暴雨量，探討所提方法之適用性。

據虞及章（1998）研究之結果，單一機率分布易被誤判為混合機率分布而導致其濫用，其中又以常態分布資料經群數判斷後改善最多，故與判斷準則之比較主要針對混合常態分布與常態分布所產生之資料，於參數檢定方面則考慮所有合成資料。

2.6.1 初步分析

實測資料之實際分布為未知，一般是以資料分別與擬合分布之設計水文量間的差異，來判定何者為最佳分布。即以大量水文站觀測值，與各種機率分布、參數推估法所推定之頻率曲線間，取 MSE 較小（即有最佳曲線配合）者為選擇標準。然，能密合大多數位於曲線尾端觀測值之分布，並不一定對曲線前端之水文量有較佳的擬合能力，最重要的是能夠掌握觀測值之統計特性。而以 MSE 判定何者為較適用之分布時，僅在於其值大小的比較；

MSEs1：資料擬合單一機率分布後之推估值與樣本值之 MSE

MSEs2：資料擬合混合機率分布後之推估值與樣本值之 MSE

若， MSEs1 < MSEs2 則表示資料以單一機率分布擬合較接近樣本值

MSEs1 > MSEs2 則表示資料以混合機率分布擬合較接近樣本值

於 MSE 值的比較上差異並不大時，在資料屬混合分布可能性不高與模式精簡的原則下，單一機率分布的採用並無不當，故加以 F 檢定判斷 MSE

之差異是否達顯著，使 MSE 之判斷更為客觀。

在混合機率分布之群數判斷上，雖可將特性明顯偏向混合機率分布（有顯著兩群）之資料篩選出，但若資料本身離散程度並不高時，反易判斷為單一機率分布，低估混合機率分布之群數（Böhning 等，1994）。故取不同長度與離散度之資料進行分析，以更加了解群數判斷準則概似比率檢定之適用範圍。

考慮由混合機率分布之參數及其檢定進行判斷，首先以單一機率分布資料擬合混合機率分布，觀察其參數特性，復利用參數檢定方法進行判斷。以擬合混合常態分布為例，則可得以下狀況時，表混合常態分布趨近於單一機率分布，資料應以單一機率分布模擬即可：

(1) 擬合 case1 得兩組成分布之平均值 μ_1、μ_2，通過雙母體平均值 $\mu_1=\mu_2$ 之檢定。

(2) 擬合 case2 得兩組成分布之變異數 ${\sigma_1}^2$、${\sigma_2}^2$，通過雙母體變異數 ${\sigma_1}^2={\sigma_2}^2$ 之檢定。

(3) 擬合 case3 得兩組成分布之平均值 μ_1、μ_2，變異數 ${\sigma_1}^2$、${\sigma_2}^2$，通過雙母體平均值 $\mu_1=\mu_2$ 與雙母體變異數 ${\sigma_1}^2={\sigma_2}^2$ 之檢定。

(4) 擬合混合常態分布得權重因子 p，當 p 或 $(1-p)$ 極小（趨近於 0）時。

其中，p 或 $(1-p)$ 是否為極小，本研究將由合成資料之分析比較討論之。

2.6.2 本研究所提方法

將資料擬合混合機率分布（僅考慮 case1、caase3、case4 及 case6），得權重因子 p 及平均值 μ_1、μ_2，則資料平均值 μ_x 與 μ_1、μ_2 之關係如式(2.6.2)所示（Day，1969）：

$$\mu_x = p\mu_1 + (1-p)\mu_2 \tag{2.6.2}$$

吾人所關注者為資料是否具有明顯混合機率分布之特性，今以分布之參數進行判斷，故針對μ_1、μ_2不相等以及 p 或$(1-p)$無顯著趨近於 0 之情況做探討。無論資料所屬分布為混合機率分布或單一機率分布，由於μ_x為μ_1、μ_2之權重相加，則μ_x對μ_1、μ_2必存在差距。反之，若μ_x與μ_1、μ_2其中之一相近並通過雙母體平均值檢定，已知 p 或$(1-p)$無顯著趨近於 0（$\neq 0$），由式(2.6.2)：

(1) $\mu_x = \mu_1$

$\mu_x = p\mu_1 + (1-p)\mu_2$ 移項 $\Rightarrow$ $\mu_x - p\mu_1 = (1-p)\mu_2$

$\geq$ $\mu_x = \mu_1$ 〃 μ_1以μ_x代入

$\Rightarrow$ $\mu_x - p\mu_x = (1-p)\mu_x = (1-p)\mu_2$ 又，$(1-p) \neq 0$

$\Rightarrow$ $\mu_x = \mu_2$

(2) $\mu_x = \mu_2$

$\mu_x = p\mu_1 + (1-p)\mu_2$ 移項 $\Rightarrow$ $\mu_x - (1-p)\mu_2 = p\mu_1$

$\geq$ $\mu_x = \mu_2$ 〃 μ_2以μ_x代入

$\Rightarrow$ $\mu_x - (1-p)\mu_x = \mu_x - \mu_x + p\mu_x = p\mu_1$ 又，$p \neq 0$

$\Rightarrow$ $\mu_x = \mu_1$

可知資料平均值μ_x與混合機率分布中兩組成分布之平均值μ_1、μ_2皆相等，資料以單一機率分布模擬即可。故本研究對μ_x、μ_1、μ_2進行$\mu_x = \mu_1$與$\mu_x = \mu_2$之雙母體平均值檢定，以為資料所屬為單一或混合機率分布之判別方法，本研究方法分析資料之步驟以流程圖說明之，如圖 2 所示。

第三章　本研究所採用之資料

本研究方法所採用之資料，計有合成資料及實測資料兩部份。合成資料又分單一機率分布與混合機率分布兩類，用於測試本研究所提之檢驗方法，實測資料則用以驗證本方法之有效性，並供實際應用時之參考，茲分述如下。

3.1 合成資料

3.1.1 單一機率分布

共計採用常態分布、對數常態分布、極端值Ⅰ型分布以及皮爾遜Ⅲ型分布等四種，樣本大小 n=30（10）100（50）150，每種樣本大小均產生 100 組，其基本統計特性如表 1 所示。

3.1.2 混合機率分布

採用混合常態分布（case1、case2、case3）產生，於組成分布考慮三種不同離散程度，權重因子 p=0.1（0.1）0.5，樣本大小 n＝30（10）100（50）150，每種樣本大小產生 100 組，其基本統計特性如表 2 所示。

3.2 實測資料

本研究採用台灣地區資料記錄年至少 30 年之年一日、二日及三日最大暴雨量資料，依地域特性分為北、中、南、東等四區，其中北區 39 站、中區 86 站、南區 106 站、東區 19 站，全省共計 250 站，關於各站之基本資料如表 3 至表 6 所示。