

## 第一章 研究背景與動機

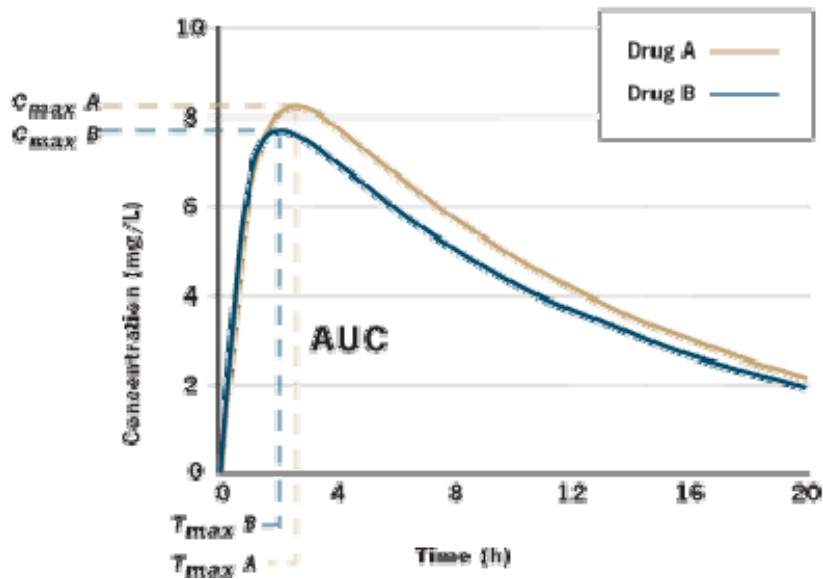
原廠專利藥上市一段時間，特別在其專利權限到期之後，為了使患者能享受較為低廉的藥物治療，美國食品藥物管理局（Food and Drug Administration，USA簡稱美國FDA）就會開放學名藥的申請，其中學名藥即是合格藥廠所生產具有與原廠專利藥相同化學成分的藥物，為了確認這種替代性藥物其療效與原廠專利藥相近，在其上市前必須通過與專利藥生體相等性的檢定（bioequivalence test）。生體相等性意指二個藥劑相等品或藥劑替代品在人體中的生體可用率（bioavailability）並無不同。其中生體可用率為藥物進入人體內之後，經由血液循環、藥物傳輸的速率及程度。

為進行生體相等性檢定，一般是在兩期雙序列交叉設計之（2×2 crossover design）之下針對自願參加的健康成人進行實驗。在此一交叉設計(表一)之下，

表1 兩期雙序列交叉設計

序列	第一期	第二期
1	原廠藥(R) 資料： $Y_{i11}, \dots, Y_{n_111}$	學名藥(T) 資料： $Y_{i21}, \dots, Y_{n_121}$
2	學名藥(T) 資料： $Y_{i12}, \dots, Y_{n_212}$	原廠藥(R) 資料： $Y_{i22}, \dots, Y_{n_222}$

首先將實驗中的健康成人被隨機地分成二個序列，其中一個序列在第一期接受原廠藥（reference drug），然後經過一段洗刷期(washout-period)後，在第二期接受學名藥。另一個序列的健康成人則是在第一期接受學名藥而在第二期接受原廠藥。每一個健康的成人皆在服用藥物後一段時間內，重複抽血測量其血液中的藥物濃度。這些在不同時間點測量得到的藥物血中濃度即稱為藥物濃度時間曲線（plasma concentration-time curve），如圖一所示，



圖一 藥物濃度時間曲線

在這測量到的藥物濃度當中最高的藥物濃度 (maximum plasma concentration, 簡稱為  $C_{max}$ ) 和達到最高濃度所需要的時間 (time until maximum plasma concentration, 簡稱為  $T_{max}$ ) 則與藥物在人體內的吸收速率有關。

一般而言，藥物濃度時間曲線下的面積 (area under the curve, 簡稱為 AUC) 為藥物在人體內的總濃度，主要為反應生體可用率。每一位受測者的藥物 AUC 可以根據 Yeh 和 Kwan (1978) 提出的梯型法加以估計，如下述

令  $C_0, C_1, \dots, C_k$  為在  $0, t_1, \dots, t_k$  個時間點所測量的藥物血中濃度，從 0 到第  $t_k$  個時間點所測量的 AUC 定義為  $AUC(0-t_k)$ ，其計算方式為

$$AUC(0 - t_k) = \sum_{i=2}^k \left( \frac{C_{i-1} + C_i}{2} \right) (t_i - t_{i-1})$$

Rowland 和 Tozer(1980)提出估計  $AUC(0\sim\infty)$  為

$$AUC(0 - \infty) = AUC(0 - t_k) + \frac{C_k}{\lambda}$$

其中  $C_k$  為測量的最後一個藥物血中濃度， $\lambda$  為代謝速率常數。

Draper和Smith(1981)則考慮簡單線性迴歸模式為

$$\log_{10}(C_i) = \beta_0 + \beta_1 t_i$$

得到 $\lambda$ 的估計為 $\hat{\lambda} = -2.303 \times \hat{\beta}_1$

令 $AUC_T$ 與 $AUC_R$ 分別代表學名藥和原廠藥的藥效，根據美國 FDA (2001)規定 $AUC_T$ 與 $AUC_R$ 的比值必須介於 4/5 到 5/4 之內，亦即

$$0.8 < \frac{AUC_T}{AUC_R} < 1.25, \text{ 取對數後得到 } -0.223 < \log AUC_T - \log AUC_R < 0.223,$$

生體相等性檢定的虛無假設和對立假設為

$$H_0: \log AUC_T - \log AUC_R \leq -0.223 \text{ 或 } \log AUC_T - \log AUC_R \geq 0.223$$

$$H_a: -0.223 < \log AUC_T - \log AUC_R < 0.223$$

上述假設可以寫成兩個統計假設：

$$H_0^1: \log AUC_T - \log AUC_R \leq -0.223$$

$$H_a^1: \log AUC_T - \log AUC_R > -0.223$$

和

$$H_0^2: \log AUC_T - \log AUC_R \geq 0.223$$

$$H_a^2: \log AUC_T - \log AUC_R < 0.223$$

傳統是分別針對這兩個單邊檢定進行顯著水準為  $\alpha$  的單邊檢定。

美國FDA (2001) 規定建立 $\log AUC_T - \log AUC_R$ 的 $100(1-2\alpha)\%$ 信賴

區間，若此一信賴區間在 $-0.223$ 至 $0.223$ 範圍之內，則宣稱兩種藥物具有生體相等性。

因為估計的 AUC 的抽樣分布為右偏分布，傳統上以對數常態分布(log normal distribution)加以描述，然而實務上的 AUC 估計式經常不是對數常態分布，因此本文考慮 Cox et al. (2007)重新定義的廣義伽瑪分布(Generalized gamma distribution)來描述 AUC 估計式的分布，廣義伽瑪分布為一個具有三參數的分布族，包含的分布中常用的有伽瑪分布(Gamma distribution)、韋伯分布(Weibull distribution)和逆伽瑪分布(Inverse Gamma distribution)也包含對數常態分布(log normal distribution)，並且根據李念純(2011)對本文的資料做適合度檢定，因此鑒於實務上的資料分析所需之下，考慮在兩期雙序列交叉設計的模式中，估計的 AUC 服從逆伽瑪分布時，進行兩種藥物的生體相等性檢定。

本文第二章介紹傳統兩期雙序列交叉設計之下，對於估計 AUC 之模式並介紹求得參數估計值的演算法。第三章提出估計 AUC 為逆伽瑪分布之下的模式，討論模式中參數的估計值，並且建立生體相等性檢定。第四章利用模擬方法在樣本數為 16 和 24 之下，比較兩個模式在三種不同參數設定下，兩者在進行生物相等性檢定的型 I 誤差

率及檢定力的表現。第五章為利用一例說明本文所提出的模式來進行生物相等性檢定，最後在第六章綜合以上研究做結語並討論未來可能研究方向。

## 第二章文獻回顧

### 2.1 兩期雙序列交叉設計實驗

考慮傳統兩期雙序列交叉設計，模式如下：

$$\log Y_{ijk} = \mu + F_{(j,k)} + P_j + S_{ik} + \log e_{ijk}, i = 1, \dots, n_k, j = k = 1, 2 \quad (2.1)$$

其中 $Y_{ijk}$ 代表第 $k$ 個順序中，在第 $j$ 個時期之下，第 $i$ 位受試者的AUC樣本觀測值。 $\mu$ 代表總平均。 $F_{(j,k)}$ 代表第 $k$ 個順序中，在第 $j$ 個時期之下藥物的固定效應，且 $F_R + F_T = 0$ 。 $P_j$ 代表第 $j$ 個時期的固定效應，且 $P_1 + P_2 = 0$ 。 $S_{ik}$ 代表代表第 $k$ 個順序中的第 $i$ 位受試者的隨機效應項，稱為受測者間的變異。 $e_{ijk}$ 代表觀測值 $Y_{ijk}$ 的隨機誤差項，稱為受測者內的變異。

模式有3個假設，即為 $S_{ik}$ 彼此獨立且服從常態分布，期望值為0，變異數為 $\sigma_S^2$ ； $e_{ijk}$ 彼此獨立且服從對數常態分布，期望值為0，變異數為 $\sigma_e^2$ ； $S_{ik}$ 與 $e_{ijk}$ 彼此獨立。由於隨機效應項與誤差項皆為常態分布且兩者獨立，相加後還是常態分布，因此在常態分布的假設下可以推導出原廠藥和學名藥藥效差異的點估計量。

令 $d_{ik} = \frac{1}{2}(Y_{i2k} - Y_{i1k})$ 為每一個順序中，第一期和第二期的AUC樣本

觀測值的差異除以且計算其平均數為  $\bar{d}_{\cdot k} = \frac{1}{n_k} \sum_{i=1}^{n_k} d_{ik}$

所以兩種藥效差異的點估計量為

$$\hat{F} = \bar{d}_{\cdot 1} - \bar{d}_{\cdot 2} = \frac{1}{2} [(\bar{Y}_{\cdot 21} - \bar{Y}_{\cdot 11}) - (\bar{Y}_{\cdot 22} - \bar{Y}_{\cdot 12})] = \bar{Y}_T - \bar{Y}_R$$

其中  $\bar{Y}_T = \frac{1}{2}(\bar{Y}_{\cdot 21} + \bar{Y}_{\cdot 12})$  和  $\bar{Y}_R = \frac{1}{2}(\bar{Y}_{\cdot 11} + \bar{Y}_{\cdot 22})$

點估計量的變異數為

$$\widehat{\text{Var}}(\hat{F}) = \hat{\sigma}_d^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)$$

$$\text{其中 } \hat{\sigma}_d^2 = \frac{1}{n_1 + n_2 - 2} \sum_{k=1}^2 \sum_{i=1}^{n_k} (d_{ik} - \bar{d}_{\cdot k})^2$$

根據(2.1)式建立  $\log\text{AUC}_T - \log\text{AUC}_R$  的  $100(1-2\alpha)\%$  信賴區間為 (L,U)

其中  $t(\alpha, n_1 + n_2 - 2)$  為  $t$  分布的第  $\alpha$  百分位數，自由度為  $n_1 + n_2 - 2$ 。

$$L = (\bar{Y}_T - \bar{Y}_R) - t(\alpha, n_1 + n_2 - 2) \hat{\sigma}_d \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$U = (\bar{Y}_T - \bar{Y}_R) + t(\alpha, n_1 + n_2 - 2) \hat{\sigma}_d \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

當此一信賴區間在  $-0.223$  至  $0.223$  範圍之內，則宣稱兩種藥物具有生體相等性。

## 2.2 求解參數估計值之演算法

在本文的研究中，考慮誤差項為逆伽瑪分布，進行兩期雙序列交



又設計模式(2.1)式，並建立聯合機率密度函數及概似函數，由於觀測到資料只有樣本 AUC，而隨機效應是無法觀測到的資訊，因此在估計參數時，必須將隨機效應 $S_{ik}$ 視為遺失值(missing value)，針對有遺失值的資料稱之為不完整資料(incomplete data)，可使用期望值-極大化演算法(Expectation maximization algorithm，簡稱 EM 演算法)求得參數估計值。

EM演算法是由Dempster et al.(1977)所提出的，主要有二步驟，稱為E步驟和M步驟。其演算步驟如下

E步驟:令 $\log L_c(y, s; \theta) = \log f(y, s; \theta)$ 為完整資料(complete data)的對數概似函數

由起始值 $\hat{\theta}^{(0)}$ 開始，計算Q函數(2.4)式

$$\begin{aligned} Q(\theta|\hat{\theta}^{(r)}) &= E\{\log f(y, s; \theta)|y, \hat{\theta}^{(r)}\} \\ &= \int \log f(y, s; \theta) f(s|y; \hat{\theta}^{(r)}) ds \end{aligned} \quad (2.2)$$

其中 $f(s|y; \hat{\theta}^{(r)})$ 為給定觀測資料和第r次參數估計值下的條件機率密度函數。

在E步驟中計算Q函數其定義為給定於觀測的資料與當前的參數估計值之下，完整資料的期望對數概似函數(expected log-likelihood function)

M步驟:目的為更新當前的參數估計值

$$\hat{\theta}^{(r+1)} = \arg \max_{\theta} Q(\theta | \hat{\theta}^{(r)}) \quad (2.3)$$

極大化Q函數獲得第r+1次參數 $\theta$ 的最大概似估計量(MLE)。

在給定起始值 $\hat{\theta}^{(0)}$ 後，反覆疊代E步驟與M步驟一直到參數估計值收斂為止。

在很多時候EM演算法E步驟的Q函數是無明確解(no closed form)，所以Wei and Tanner (1990)提出以蒙地卡羅EM演算法(Monte Carlo EM algorithm，簡稱MCEM演算法)，來近似EM演算法E步驟的Q函數。如下所述

E步驟:

$$\tilde{Q}(\theta | \hat{\theta}^{(r)}) = \frac{1}{m} \sum_{t=1}^m \log f(y, s_t^{(r)}; \theta)$$

其中 $s_1^{(r)}, s_2^{(r)}, \dots, s_m^{(r)}$ 為生成m個隨機樣本來自條件機率密度函數 $f(s|y; \hat{\theta}^{(r)})$ ，且在隨機樣本數夠大的情況下，根據大數法則(the law of large numbers)， $\tilde{Q}(\theta | \hat{\theta}^{(r)})$ 將可合理的近似到EM演算法的(2.2)式。

M步驟做法和EM演算法是相同的。

因為MCEM演算法需要生成大量的隨機樣本，相當耗費計算成本，也導致收斂速度緩慢，所以Delyon et al. (1999)提出SAEM演算法(stochastic approximation expectation maximization algorithm，簡稱SAEM 演算法)，其做法是使用一個遞減的權重序列稱為平滑參數(smoothing parameter)，這個方法可以確保較早期的疊代的資訊被慢

慢捨棄和重點放在比較近期的疊代，SAEM演算法是以一種隨機逼近的過程來取代MCEM演算法的E步驟並且在這隨機近似的過程可以生成的少量的隨機樣本，大幅的提升計算速度。SAEM演算法步驟如下E步驟：

生成 $s_1, s_2, \dots, s_m$ 隨機樣本來自條件機率密度函數 $f(s|y; \hat{\theta}^{(r)})$ ，其中 $r$ 為第 $r$ 次的疊代。

$$\tilde{Q}(\theta|\hat{\theta}^{(r)}) = \tilde{Q}(\theta|\hat{\theta}^{(r-1)}) + \delta_k \left[ \frac{1}{m} \sum_{t=1}^m \log f(y, s_t^{(r)}; \theta) - \tilde{Q}(\theta|\hat{\theta}^{(r-1)}) \right]$$

平滑參數 $\delta_r$ 是恆正且遞減的序列需滿足 $\sum_{r=1}^{\infty} \delta_r = \infty$ ， $\sum_{r=1}^{\infty} \delta_r^2 < \infty$

根據Meza et al.(2010)的研究，生成隨機樣本數量為10及平滑參數形式為

$$\delta_r = \begin{cases} 1 & , \quad 1 \leq r \leq r_s \\ \frac{1}{r - r_s} & , \quad r_s + 1 \leq r \leq r_T \end{cases}$$

其中 $r$ 的定義為第 $r$ 次的疊代， $r_s$ 為平滑開始的疊代次數， $r_T$ 為總疊代次數。根據Meza et al.(2010)指出在執行SAEM演算法時，首先必須先固定疊代的總次數和平滑步驟開始之前的迭代次數，決定這兩個疊代次數可以透過圖形來監控Q函數在每次疊代過程中其遞增的情況，當Q函數值穩定後即可開始平滑步驟，因此這也代表必須執行二次SAEM

演算法來求得參數估計值，首先第一次執行目的為找尋開始平滑步驟的疊代次數，例如當執行一次SAEM演算法時，前20次疊代中Q函數值已經趨於穩定不再增加， $r_s$ 可設為20次，平滑參數可設定為

$$\delta_r = \begin{cases} 1 & , 1 \leq r \leq 20 \\ \frac{1}{r-20} & , 21 \leq r \leq 40 \end{cases}$$

## 2.3 估計參數標準差

隨機效應變數S在給定觀測資料和參數下條件機率密度函數為

$$f(s|y; \boldsymbol{\theta}) = \frac{f(y, s; \boldsymbol{\theta})}{f(y; \boldsymbol{\theta})}$$

取對數後為 $\log f(y; \boldsymbol{\theta}) = \log f(y, s; \boldsymbol{\theta}) - \log f(s|y; \boldsymbol{\theta})$ ，因此不完整資料的對數概似函數可以表達成

$$\log L_o(\boldsymbol{\theta}; y) = \log L_c(\boldsymbol{\theta}; y, s) - \log f(s|y; \boldsymbol{\theta})$$

在等號兩邊對參數做二次微分後取負號為

$$H_o(\boldsymbol{\theta}; y) = H_c(\boldsymbol{\theta}; y, s) + \frac{\partial^2}{\partial \theta \partial \theta'} \log f(s|y; \boldsymbol{\theta})$$

取期望值後，不完整資料的訊息矩陣(information matrix)為

$$H_o(\boldsymbol{\theta}; y) = H_c(\boldsymbol{\theta}; y, s) - H_m(\boldsymbol{\theta}; y) \quad (2.4)$$

其中 $H_c(\boldsymbol{\theta}; y) = -E \left\{ \frac{\partial^2}{\partial \theta \partial \theta'} \log f(y, s; \boldsymbol{\theta}) | y; \boldsymbol{\theta} \right\}$

$$H_m(\boldsymbol{\theta}; y) = -E \left\{ \frac{\partial^2}{\partial \theta \partial \theta'} \log f(s|y; \boldsymbol{\theta}) | y; \boldsymbol{\theta} \right\}$$

也就是不完整觀測資料的訊息矩陣為完整資料的訊息矩陣減去遺失

資料的訊息矩陣。

Louis(1982)證明遺失資料訊息矩陣可以表達成

$$\begin{aligned} H_m(\boldsymbol{\theta}) &= -E \left[ \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \log f(s|y; \boldsymbol{\theta}) \mid y; \boldsymbol{\theta} \right] \\ &= E \left[ \left( \frac{\partial}{\partial \boldsymbol{\theta}} \log f(y, s; \boldsymbol{\theta}) \right) \left( \frac{\partial}{\partial \boldsymbol{\theta}} \log f(y, s; \boldsymbol{\theta}) \right)' \mid y; \boldsymbol{\theta} \right] - U_o(\boldsymbol{\theta}; y) U_o'(\boldsymbol{\theta}; y) \end{aligned}$$

其中不完整資料的對數概似函數的分數函數(score function)為

$$U_o(\boldsymbol{\theta}) = E_{\theta} [U_c(\boldsymbol{\theta}; y, s) \mid y; \boldsymbol{\theta}] = E \left[ \left( \frac{\partial}{\partial \boldsymbol{\theta}} \log f(y, s; \boldsymbol{\theta}) \right) \mid y; \boldsymbol{\theta} \right]$$

因此將(2.4)式重新表達，不完整資料的訊息矩陣為

$$\begin{aligned} H_o(\boldsymbol{\theta}) &= H_c(\boldsymbol{\theta}) - H_m(\boldsymbol{\theta}) \\ &= E_{\theta} [H_c(\boldsymbol{\theta}; y, s) \mid y; \boldsymbol{\theta}] - E_{\theta} [U_c(\boldsymbol{\theta}; y, s) U_c'(\boldsymbol{\theta}; y, s) \mid y; \boldsymbol{\theta}] + U_o(\boldsymbol{\theta}; y) U_o'(\boldsymbol{\theta}; y) \end{aligned} \quad (2.5)$$

其中  $H_c(\boldsymbol{\theta}) = -\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \log f(y, s; \boldsymbol{\theta})$  為完整資料的對數概似函數對參數

做二次偏微分取負號。

$U_c(\boldsymbol{\theta}; y, s) = \frac{\partial}{\partial \boldsymbol{\theta}} \log f(\boldsymbol{\theta}; y, s)$  為完整資料的對數概似函數的分數函數

此外根據蒙地卡羅方法來近似(2.5)式如下

$$\begin{aligned} E_{\theta} [H_c(\boldsymbol{\theta}; y, s) \mid y] &\approx \frac{1}{m} \sum_{t=1}^m \left\{ H_c \left( \hat{\boldsymbol{\theta}}^{(r)}; y, s_t^{(r)} \right) \right\} \\ E_{\theta} [U_c(\boldsymbol{\theta}; y, s) U_c'(\boldsymbol{\theta}; y, s) \mid y] &\approx \frac{1}{m} \sum_{t=1}^m \left\{ U_c \left( \hat{\boldsymbol{\theta}}^{(r)}; y, s_t^{(r)} \right) U_c' \left( \hat{\boldsymbol{\theta}}^{(r)}; y, s_t^{(r)} \right) \right\} \\ U_o(\boldsymbol{\theta}; y) = E_{\theta} \{ U_c(\boldsymbol{\theta}; y, s) \mid y; \boldsymbol{\theta} \} &\approx \frac{1}{m} \sum_{t=1}^m U_c \left( \hat{\boldsymbol{\theta}}^{(r)}; y, s_t^{(r)} \right) \end{aligned} \quad (2.6)$$

將上式運用到 SAEM 演算法如下

$$\begin{aligned}
 J_r &= J_{r-1} + \\
 &\delta_r \left( \frac{1}{m} \sum_{t=1}^m \left\{ H_c(\hat{\theta}^{(r)}; y, s_t^{(r)}) - U_c(\hat{\theta}^{(r)}; y, s_t^{(r)}) U_c'(\hat{\theta}^{(r)}; y, s_t^{(r)}) \right\} - J_{r-1} \right) \\
 g_r &= J_{r-1} + \delta_r \left( \frac{1}{m} \sum_{t=1}^m U_c(\hat{\theta}^{(r)}; y, s_t^{(r)}) - g_{r-1} \right) \quad (2.7)
 \end{aligned}$$

$J_r$ 為(2.6)前兩式的合併，也就是完整資料的對數概似函數對參數做二次偏微分取負號後再減去完整資料的對數概似函數的分數函數與其轉置後的乘積。 $g_r$ 為不完整資料的對數概似函數的分數函數與其轉置後的乘積。因此不完整資料的訊息矩陣(observed information matrix)為  $\hat{I}_r = J_r + g_r g_r'$  且參數的變異數共變異數矩陣為  $\hat{I}_r^{-1}$ 。

### 第三章 統計方法

本章考慮兩期雙序列交叉設計模式，將誤差項分布為逆伽瑪分布，並利用所提出的模式建立概似函數，並以 SAEM 演算法求得參數的最大概似估計量與其標準差來進行生體相等性檢定。

#### 3.1 廣義伽瑪分布

根據 Cox et al. (2007) 的結果，若隨機變數  $y$  服從廣義伽瑪分布，其機率密度函數為

$$f_{GG}(y; \beta, \sigma, \lambda) = \begin{cases} \frac{|\lambda|}{\sigma y \Gamma(\lambda^{-2})} \left[ \lambda^{-2} (e^{-\beta y})^{\frac{\lambda}{\sigma}} \right]^{\lambda^{-2}} e^{\left[ -\lambda^{-2} (e^{-\beta y})^{\frac{\lambda}{\sigma}} \right]}, & \lambda = 0 \\ \frac{1}{\sqrt{2\pi\sigma y}} \exp - \left[ \frac{(\log y - \beta)^2}{2\sigma^2} \right] & , \lambda \neq 0 \end{cases} \quad (3.1)$$

其中  $\beta$  為位置參數(location parameter)， $\sigma$  為尺度參數(scale parameter)， $\lambda$  為形狀參數(shape parameter)。根據 Gomes 等人(2008) 的研究，廣義伽瑪分布的期望值為

$$E(Y) = \frac{\lambda^{\frac{2\sigma}{\lambda}} e^{\beta} \Gamma(\lambda^{-2} + \frac{\sigma}{\lambda})}{\Gamma(\lambda^{-2})}$$

當  $\lambda=0$ ， $f_{GG}(y)$  對數常態分布； $\lambda=1$ ， $f_{GG}(y)$  為韋伯分布；

$\lambda=\sigma=1$ ， $f_{GG}(y)$  為指數分布； $\lambda=\sigma$ ， $f_{GG}(y)$  為伽瑪分布；

$\lambda = -\sigma$ ， $f_{GG}(y)$  為逆伽瑪分布

因此若隨機變數  $X$  服從逆伽瑪分布，則其機率密度函數

$$f(x; \lambda, \beta) = \frac{\beta^{\lambda^{-2}}}{\Gamma(\lambda^{-2})} x^{-\lambda^{-2}-1} e^{-\frac{\beta}{x}} \quad (3.2)$$

$\lambda^{-2} > 0$  為形狀參數； $\beta > 0$  為尺度參數

期望值為： $E(X) = \frac{\beta}{\lambda^{-2}-1}$  且滿足  $\lambda^{-2} > 1$ 。

變異數為： $Var(x) = \frac{\beta^2}{(\lambda^{-2}-1)^2(\lambda^{-2}-2)}$  且滿足  $\lambda^{-2} > 2$ 。

相關的分佈：

若隨機變數  $X$  服從  $Gamma(\lambda, \beta)$ ，則  $\frac{1}{X}$  服從  $Inverse\ Gamma(\lambda, \beta)$

若隨機變數  $X$  服從  $Inverse\ Gamma(\lambda, \beta)$  且  $k$  為常數，則  $kX$  服從  $Inverse\ Gamma(\lambda, k\beta)$ 。

## 3.2 統計模式

考慮兩期雙序列交叉設計模式如下

$$Y_{ijk} = e^{\mu_{jk} + P_j + S_{ik}} \cdot \varepsilon_{ijk}, i = 1, \dots, n_k, j = k = 1, 2 \quad (3.3)$$

其中  $Y_{ijk}$  代表第  $k$  個順序，在第  $j$  個時期之下，第  $i$  位受試者的 AUC 樣本觀測值。 $\mu_{jk}$  代表藥物的平均藥效，若  $j=k$  為原廠藥，若  $j \neq k$  為學名藥。

$P_j$  代表第  $j$  個時期的固定效應，且  $P_1 + P_2 = 0$ 。 $S_{ik}$  代表第  $k$  個順序中



的第  $i$  位受試者的隨機效應項。 $\varepsilon_{ijk}$  為觀測值  $Y_{ijk}$  的隨機誤差項。模式有 3 個假設為  $S_{ik}$  彼此獨立且服從常態分布，期望值為 0，變異數為  $\sigma_S^2$ 。 $e_{ijk}$  彼此獨立且服從逆伽瑪分布， $\lambda$  為形狀參數， $\beta$  為尺度參數。 $S_{ik}$  與  $e_{ijk}$  彼此獨立。

考慮模式(3.3)在給定  $S_{ik}$  時，根據 3.1 節的相關分布， $Y_{ijk}$  給定  $S_{ik}$  時會服從逆伽瑪分布，所以其機率密度函數為

$$f(y_{ijk}|s_{ik}; \theta) = \frac{[\beta e^{\mu_{jk}+p_j+s_{ik}}]^{\lambda-2}}{\Gamma(\lambda-2)} y_{ijk}^{-\lambda-2-1} e^{-\frac{(\beta e^{\mu_{jk}+p_j+s_{ik}})}{y_{ijk}}}$$

因此  $E(Y_{ijk}|S_{ik}; \theta) = e^{\mu_{jk}+p_j+s_{ik}} E(\varepsilon_{ijk})$ ，考慮當  $E(\varepsilon_{ijk}) = 1$ ，則

$\beta = \lambda^{-2} - 1$ ，因此機率密度函數可以改寫成

$$f(y_{ijk}|s_{ik}; \theta) = \frac{[(\lambda^{-2} - 1)e^{\mu_{jk}+p_j+s_{ik}}]^{\lambda-2}}{\Gamma(\lambda-2)} y_{ijk}^{-\lambda-2-1} e^{-\frac{(\lambda^{-2}-1)e^{\mu_{jk}+p_j+s_{ik}}}{y_{ijk}}}$$

(3.4)

因為  $(Y_{i1k}, Y_{i2k})'$  為第  $i$  位受測者在第  $k$  個順序，為成對的 AUC 樣本觀測值，因此在給定隨機效應  $S_{ik}$  下，兩者可視為獨立，當  $k=1$  (第 1 個序列)，其聯合機率密度函數為

$$\begin{aligned} f(y_{i11}, y_{i21}, s_{i1}; \theta) &= f(y_{i11}, y_{i21} | s_{i1}; \theta) \cdot f(s_{i1}; \theta) \\ &= f(y_{i11} | s_{i1}; \theta) \cdot f(y_{i21} | s_{i1}; \theta) \cdot f(s_{i1}; \theta) \\ &= \frac{[(\lambda^{-2} - 1)e^{\mu_{R}+p_1+s_{i1}}]^{\lambda-2}}{\Gamma(\lambda-2)} y_{i11}^{-\lambda-2-1} e^{-\frac{[(\lambda^{-2}-1)e^{\mu_{R}+p_1+s_{i1}}]}{y_{i11}}} \\ &\quad \cdot \frac{[(\lambda^{-2} - 1)e^{\mu_{T}+p_2+s_{i1}}]^{\lambda-2}}{\Gamma(\lambda-2)} y_{i21}^{-\lambda-2-1} e^{-\frac{[(\lambda^{-2}-1)e^{\mu_{T}+p_2+s_{i1}}]}{y_{i21}}} \cdot \frac{1}{\sqrt{2\pi}\sigma_S} e^{-\frac{s_{i1}^2}{2\sigma_S^2}} \end{aligned}$$

且第一個順序和第二個順序的 AUC 樣本觀測值彼此是獨立的，所以概似函數為

$$\begin{aligned}
L(\boldsymbol{\theta}|y_{i1k}, y_{i2k}, s_{ik}) &= \prod_{k=1}^2 \prod_{i=1}^{n_k} f(y_{i1k}, y_{i2k}, s_{ik}; \boldsymbol{\theta}) \\
&= \prod_{k=1}^2 \prod_{i=1}^{n_k} f(y_{i1k}|s_{ik}; \boldsymbol{\theta}) \cdot f(y_{i2k}|s_{ik}; \boldsymbol{\theta}) \cdot f(s_{ik}; \boldsymbol{\theta}) \\
&= \prod_{i=1}^{n_1} \frac{[(\lambda^{-2} - 1)e^{\mu_R+p_1+s_{i1}}]^{\lambda^{-2}}}{\Gamma(\lambda^{-2})} y_{i11}^{-\lambda^{-2}-1} e^{-\frac{[(\lambda^{-2}-1) e^{\mu_R+p_1+s_{i1}}]}{y_{i11}}} \\
&\quad \cdot \frac{[(\lambda^{-2} - 1)e^{\mu_T+p_2+s_{i1}}]^{\lambda^{-2}}}{\Gamma(\lambda^{-2})} y_{i21}^{-\lambda^{-2}-1} e^{-\frac{[(\lambda^{-2}-1) e^{\mu_T+p_2+s_{i1}}]}{y_{i21}}} \cdot \frac{1}{\sqrt{2\pi}\sigma_S} e^{-\frac{s_{i1}^2}{2\sigma_S^2}} \\
&\quad \prod_{i=1}^{n_2} \frac{[(\lambda^{-2} - 1)e^{\mu_T+p_1+s_{i2}}]^{\lambda^{-2}}}{\Gamma(\lambda^{-2})} y_{i12}^{-\lambda^{-2}-1} e^{-\frac{[(\lambda^{-2}-1) e^{\mu_T+p_1+s_{i2}}]}{y_{i12}}} \\
&\quad \cdot \frac{[(\lambda^{-2} - 1)e^{\mu_R+p_2+s_{i2}}]^{\lambda^{-2}}}{\Gamma(\lambda^{-2})} y_{i22}^{-\lambda^{-2}-1} e^{-\frac{[(\lambda^{-2}-1) e^{\mu_R+p_2+s_{i2}}]}{y_{i22}}} \frac{1}{\sqrt{2\pi}\sigma_S} e^{-\frac{s_{i2}^2}{2\sigma_S^2}}
\end{aligned} \tag{3.6}$$

其中參數  $\boldsymbol{\theta} = (\mu_R, \mu_T, p, \lambda, \sigma_S)^T$ 。

由(3.6)式可以得知，在概似函數中觀測到資料只有樣本AUC，而隨機效應是無法觀測到的資訊，稱為不完整資料的概似函數，定義為

$L(\boldsymbol{\theta}|y_{i1k}, y_{i2k})$ ，所以不完整資料和完整資料的關係為：

$$\begin{aligned}
L(\boldsymbol{\theta}|y_{i1k}, y_{i2k}) &= \prod_{k=1}^2 \prod_{i=1}^{n_k} f(y_{i1k}, y_{i2k}, s_{ik}; \boldsymbol{\theta}) \\
&= \int_{-\infty}^{\infty} \prod_{k=1}^2 \prod_{i=1}^{n_k} f(y_{i1k}, y_{i2k}, s_{ik}; \boldsymbol{\theta}) ds_{ik}
\end{aligned}$$

因此根據EM演算法(2.4)式，Q函數可表達為：

$$\begin{aligned}
Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(r)}) &= E\{\log L_c(\boldsymbol{\theta}|y_{i1k}, y_{i2k}, s_{ik})|y_{i1k}, y_{i2k}, \hat{\boldsymbol{\theta}}^{(r)}\} \\
&= \int_{-\infty}^{\infty} \sum_{k=1}^2 \sum_{i=1}^{n_k} \log f(y_{i1k}, y_{i2k}, s_{ik}; \boldsymbol{\theta}) \times f(s_{ik}|y_{i1k}, y_{i2k}; \hat{\boldsymbol{\theta}}^{(r)}) ds_{ik}
\end{aligned}$$

由上式可知Q函數無明確解(no closed-form)，所以需用蒙地卡羅法

來近似。因此MCEM演算法的E步驟為

在第r次疊代下生成  $s_1^{(r)}, s_2^{(r)}, \dots, s_m^{(r)}$  隨機樣本來自機率密度函數為

$$f(s_{ik}|y_{i1k}, y_{i2k}; \boldsymbol{\theta}) = \frac{f(y_{i1k}, y_{i2k}, s_{ik}; \boldsymbol{\theta})}{\int_{-\infty}^{\infty} f(y_{i1k}, y_{i2k}, s_{ik}; \boldsymbol{\theta}) ds_{ik}} \propto f(y_{i1k}, y_{i2k}, s_{ik}; \boldsymbol{\theta})$$

$$\begin{aligned}
\tilde{Q}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(r)}) &= \frac{1}{m} \sum_{t=1}^m \log L_c(\boldsymbol{\theta}|y_{i1k}, y_{i2k}, s_t^{(r)}) \\
&= \frac{1}{m} \sum_{t=1}^m \left\{ \sum_{k=1}^2 \sum_{i=1}^{n_k} \log f(y_{i1k}, y_{i2k}, s_{ik}^{(r)}; \boldsymbol{\theta}) \right\} \\
&= \frac{1}{m} \sum_{t=1}^m \sum_{i=1}^n \left\{ \log f(y_{i11}|s_{i1t}^{(r)}; \boldsymbol{\theta}) + \log f(y_{i21}|s_{i1t}^{(r)}; \boldsymbol{\theta}) + \log f(s_{i1t}^{(r)}; \boldsymbol{\theta}) \right. \\
&\quad \left. + \log f(y_{i12}|s_{i2t}^{(r)}; \boldsymbol{\theta}) + \log f(y_{i22}|s_{i2t}^{(r)}; \boldsymbol{\theta}) + \log f(s_{i2t}^{(r)}; \boldsymbol{\theta}) \right\} \quad (3.7)
\end{aligned}$$

其中  $k=1,2$  ,  $n_1 = n_2 = n$

### 3.3 SAEM 演算法

前一節中已經介紹在兩期雙序列交叉設計模式的概似函數以及 MCEM 演算法的 Q 函數，因此運用到 SAEM 演算法如下

E 步驟:

由起始值  $\hat{\boldsymbol{\theta}}^{(0)}$  開始

在第  $r$  次疊代下生成  $s_1^{(r)}, s_2^{(r)}, \dots, s_m^{(r)}$  隨機樣本來自機率密度函數為

$$f(s_{ik} | y_{i1k}, y_{i2k}; \hat{\boldsymbol{\theta}}^{(r)}) \\ = \prod_{k=1}^2 \prod_{i=1}^{n_k} f(y_{i1k} | s_{ik}; \hat{\boldsymbol{\theta}}^{(r)}) \cdot f(y_{i2k} | s_{ik}; \hat{\boldsymbol{\theta}}^{(r)}) \cdot f(s_{ik}; \hat{\boldsymbol{\theta}}^{(r)}) \quad (3.8)$$

其中  $r$  為第  $r$  次的疊代， $\hat{\boldsymbol{\theta}}^{(r)}$  為第  $r$  次的疊代的參數估計值。

隨機近似過程: 計算 Q 函數

$$\tilde{Q}(\boldsymbol{\theta} | \hat{\boldsymbol{\theta}}^{(r)}) = \tilde{Q}(\boldsymbol{\theta} | \hat{\boldsymbol{\theta}}^{(r-1)}) + \delta_r \left[ \frac{1}{m} \sum_{t=1}^m \log f(y_{i1k}, y_{i2k}, s_t^{(r)}; \boldsymbol{\theta}) - \tilde{Q}(\boldsymbol{\theta} | \hat{\boldsymbol{\theta}}^{(r-1)}) \right] \quad (3.9)$$

其中

$$\log f(y_{i1k}, y_{i2k}, s_t^{(r)}; \boldsymbol{\theta}) \\ = \sum_{i=1}^n \log f(y_{i11} | s_{i1t}^{(r)}; \boldsymbol{\theta}) + \log f(y_{i21} | s_{i1t}^{(r)}; \boldsymbol{\theta}) + \log f(s_{i1t}^{(r)}; \boldsymbol{\theta}) \\ + \log f(y_{i12} | s_{i2t}^{(r)}; \boldsymbol{\theta}) + \log f(y_{i22} | s_{i2t}^{(r)}; \boldsymbol{\theta}) + \log f(s_{i2t}^{(r)}; \boldsymbol{\theta})$$

平滑參數 $\delta_r$  是恆正且遞減的序列需滿足  $\sum_{r=1}^{\infty} \delta_r = \infty$  ,  $\sum_{r=1}^{\infty} \delta_r^2 < \infty$

$$\delta_k = \begin{cases} 1, & 1 \leq r \leq 20 \\ \frac{1}{r - r_s}, & r_s + 1 \leq r \leq r_T \end{cases}$$

$r_s$  定義為平滑(smoothing)開始的疊代次數,  $r_T$  定義為總疊代次數, 在本文的研究中, 前 20 次疊代中 Q 函數值已經趨於穩定不再增加和大幅波動, 所以 $r_s$  設為 20 次, 平滑參數設定為

$$\text{平滑參數}\delta_r = \begin{cases} 1, & 1 \leq r \leq 20 \\ \frac{1}{r - 20}, & 20 + 1 \leq r \leq 40 \end{cases}$$

M 步驟為更新當前的參數估計值

$$\hat{\theta}^{(r+1)} = \arg \max_{\theta} Q(\theta | \hat{\theta}^{(r)})$$

極大化 Q 函數所使用的是 R 軟體中的 `nlm()` 指令求得參數  $\theta$  之最大概似估計 (MLE) 即  $\hat{\theta} = (\hat{\mu}_R, \hat{\mu}_T, \hat{p}, \hat{\lambda}, \hat{\sigma}_S)$

在給定起始值  $\hat{\theta}^{(0)}$  後, 執行一次 E 步驟與 M 步驟得到  $\hat{\theta}^{(1)}$ , 反覆執行 40 次得到  $\hat{\theta}^{(40)}$  即為收斂的參數估計值。

根據 (2.7) 式 SAEM 演算法參數標準差的估計為

$$J_r = J_{r-1} + \delta_r \left\{ \frac{1}{m} \sum_{t=1}^m \left[ H_c \left( \hat{\theta}^{(r)}; y_{i1k}, y_{i2k}, s_t^{(r)} \right) - U_c \left( \hat{\theta}^{(r)}; y_{i1k}, y_{i2k}, s_t^{(r)} \right) U_c' \left( \hat{\theta}^{(r)}; y_{i1k}, y_{i2k}, s_t^{(r)} \right) \right] - J_{r-1} \right\}$$

$$\mathbf{g}_r = \mathbf{g}_{r-1} + \delta_r \left( \frac{1}{m} \sum_{t=1}^m U_c \left( \hat{\boldsymbol{\theta}}^{(r)}; y_{i1k}, y_{i2k}, s_t^{(r)} \right) - \mathbf{g}_{r-1} \right)_{\theta=\hat{\theta}} \quad (3.10)$$

其中

$$\begin{aligned} H_c \left( \hat{\boldsymbol{\theta}}^{(r)}; y_{i1k}, y_{i2k}, s_t^{(r)} \right) &= \frac{-\partial^2}{\partial \theta \partial \theta'} \log L_c \left( y_{i1k}, y_{i2k}, s_t^{(r)}; \boldsymbol{\theta} \right) \\ U_c \left( \hat{\boldsymbol{\theta}}^{(r)}; y_{i1k}, y_{i2k}, s_t^{(r)} \right) U_c' \left( \hat{\boldsymbol{\theta}}^{(r)}; y_{i1k}, y_{i2k}, s_t^{(r)} \right) \\ &= \left( \frac{\partial}{\partial \theta} \log L_c \left( y_{i1k}, y_{i2k}, s_t^{(r)}; \boldsymbol{\theta} \right) \right)_{\theta=\hat{\theta}} \left( \frac{\partial}{\partial \theta} \log L_c \left( y_{i1k}, y_{i2k}, s_t^{(r)}; \boldsymbol{\theta} \right) \right)'_{\theta=\hat{\theta}} \\ U_c \left( \hat{\boldsymbol{\theta}}^{(r)}; y_{i1k}, y_{i2k}, s_t^{(r)} \right) &= \frac{\partial}{\partial \theta} \log L_c \left( y_{i1k}, y_{i2k}, s_t^{(r)}; \boldsymbol{\theta} \right) \end{aligned}$$

$J_r$  為完整資料的對數概似函數對參數做二次偏微分取負號後，再減去

完整資料的對數概似函數的分數函數與其自身轉置後的乘積後取期

望值。 $\mathbf{g}_r$  為不完整資料的對數概似函數的分數函數與其轉置後的乘積。

最後得到的觀測的訊息矩陣為  $\hat{I}_r = J_r + \mathbf{g}_r \mathbf{g}_r'$  且變異數共變異數矩陣

為  $\hat{\Sigma} = \hat{I}_r^{-1}$ ，其中  $\hat{\Sigma}$  為

$$\begin{bmatrix} \widehat{\text{Var}}(\hat{\mu}_R) & \widehat{\text{Cov}}(\hat{\mu}_R, \hat{\mu}_T) & \widehat{\text{Cov}}(\hat{\mu}_R, \hat{p}) & \widehat{\text{Cov}}(\hat{\mu}_R, \hat{\lambda}) & \widehat{\text{Cov}}(\hat{\mu}_R, \hat{\sigma}_s) \\ \widehat{\text{Cov}}(\hat{\mu}_R, \hat{\mu}_T) & \widehat{\text{Var}}(\hat{\mu}_T) & \widehat{\text{Cov}}(\hat{\mu}_T, \hat{p}) & \widehat{\text{Cov}}(\hat{\mu}_T, \hat{\lambda}) & \widehat{\text{Cov}}(\hat{\mu}_T, \hat{\sigma}_s) \\ \widehat{\text{Cov}}(\hat{\mu}_R, \hat{p}) & \widehat{\text{Cov}}(\hat{\mu}_T, \hat{p}) & \widehat{\text{Var}}(\hat{p}) & \widehat{\text{Cov}}(\hat{p}, \hat{\lambda}) & \widehat{\text{Cov}}(\hat{p}, \hat{\sigma}_s) \\ \widehat{\text{Cov}}(\hat{\mu}_R, \hat{\lambda}) & \widehat{\text{Cov}}(\hat{\mu}_T, \hat{\lambda}) & \widehat{\text{Cov}}(\hat{p}, \hat{\lambda}) & \widehat{\text{Var}}(\hat{\lambda}) & \widehat{\text{Cov}}(\hat{\lambda}, \hat{\sigma}_s) \\ \widehat{\text{Cov}}(\hat{\mu}_R, \hat{\sigma}_s) & \widehat{\text{Cov}}(\hat{\mu}_T, \hat{\sigma}_s) & \widehat{\text{Cov}}(\hat{p}, \hat{\sigma}_s) & \widehat{\text{Cov}}(\hat{\lambda}, \hat{\sigma}_s) & \widehat{\text{Var}}(\hat{\sigma}_s) \end{bmatrix}$$

### 3.4 執行 SAEM 演算法估計參數過程

步驟 1: 設定起始值  $r=0$  ,  $\hat{\theta}^{(r)} = \{\hat{\mu}_R^{(r)}、\hat{\mu}_T^{(r)}、\hat{\rho}^{(r)}、\hat{\lambda}^{(r)}、\hat{\sigma}_s^{(r)}\}$ 。

步驟 2: 生成  $m$  個隨機樣本來自條件機率密度函數(3.8)式。

步驟 3: 計算  $Q$  函數(3.9)式並且極大化，得到參數估計值

$$\{\hat{\mu}_R、\hat{\mu}_T、\hat{\rho}_1、\hat{\rho}_2、\hat{\lambda}、\hat{\sigma}_s\}。$$

步驟 4: 利用估計的參數 $\{\hat{\mu}_R、\hat{\mu}_T、\hat{\rho}_1、\hat{\rho}_2、\hat{\lambda}、\hat{\sigma}_s\}$ 計算(3.10)式求得

估計的參數訊息矩陣。

步驟 5:  $k=k+1$  且令  $\theta^{(r+1)} = \{\hat{\mu}_R、\hat{\mu}_T、\hat{\rho}、\hat{\lambda}、\hat{\sigma}_s\}$ 。

步驟 6: 重複步驟 2 至步驟 5 直到疊代次數  $r = r_T$ 。

根據 SAEM 演算法求得參數的最大概似估計量及  $\hat{\Sigma}$  變異數共變異數矩陣，可以得到  $\hat{\mu}_R、\hat{\mu}_T、\widehat{\text{Var}}(\hat{\mu}_R)、\widehat{\text{Var}}(\hat{\mu}_T)、\widehat{\text{Cov}}(\hat{\mu}_R, \hat{\mu}_T)$ ，在正規條件下最大概似估計量有漸近常態分布的性質，建立  $\mu_T - \mu_R$  的

100(1-2 $\alpha$ )%信賴區為

$$(L, U) = \left( (\hat{\mu}_T - \hat{\mu}_R) - z_\alpha \times \sqrt{\widehat{\text{Var}}(\hat{\mu}_R) + \widehat{\text{Var}}(\hat{\mu}_T) - 2 \times \widehat{\text{Cov}}(\hat{\mu}_R, \hat{\mu}_T)}, \right. \\ \left. (\hat{\mu}_T - \hat{\mu}_R) + z_\alpha \times \sqrt{\widehat{\text{Var}}(\hat{\mu}_R) + \widehat{\text{Var}}(\hat{\mu}_T) - 2 \times \widehat{\text{Cov}}(\hat{\mu}_R, \hat{\mu}_T)} \right)$$

當建構的信賴區間(L, U)，落於-0.223 至 0.223 範圍之內，則宣稱學名藥與原廠專利藥具有生體相等性。

## 第四章 模擬研究

為瞭解本文所提非對數常態分布的生體相等性檢定表現，在本章將執行一個模擬研究。在兩期雙序列交叉設計模式下(3.3)式，誤差項服從逆伽瑪分布下模擬資料並進行生體相等性檢定，並且和傳統模式比較兩者的型 I 誤差率及檢定力的表現。

### 4.1 模擬方法

此一模擬研究中，假設二組受試者的人數 $n$ 為16和24之下的兩期雙序列交叉設計模式，考慮估計的AUC為逆伽瑪分布之下其形狀參數 $\lambda$ 分別為-0.15，-0.35，隨機效應 $S_{ik}$ 服從常態分布，平均數為0，標準差 $\sigma_s$ 為0.3和0.5，考慮設定 $\frac{AUC_T}{AUC_R}$ 的比值分別有：0.8，0.95，1，1.1，1.25，所以當設定 $\frac{AUC_T}{AUC_R}$ 比值為0.8及1.25時，生體相等性檢定結果宣告成立時，為型 I 誤差率的表現。當設定 $\frac{AUC_T}{AUC_R}$ 比值為0.95、1、1.1時，生體相等性檢定結果宣告成立時為檢定力的表現。本文的模擬研究總結如下

樣本數 $n=16$ 時

Case1: $\lambda$  為-0.15； $\sigma_s$  為 0.3； $\frac{AUC_T}{AUC_R}$  比值為 0.8、0.95，1，1.1，1.25。

Case2: $\lambda$  為-0.35； $\sigma_s$  為 0.3； $\frac{AUC_T}{AUC_R}$  比值為 0.8、0.95，1，1.1，1.25。



Case3: $\lambda$  為-0.35； $\sigma_s$  為 0.5； $\frac{AUC_T}{AUC_R}$  比值為 0.8、0.95，1，1.1，1.25。

樣本數 $n=24$ 時

Case4: $\lambda$  為-0.15； $\sigma_s$  為 0.3； $\frac{AUC_T}{AUC_R}$  比值為 0.8、0.95，1，1.1，1.25。

Case5: $\lambda$  為-0.35； $\sigma_s$  為 0.3； $\frac{AUC_T}{AUC_R}$  比值為 0.8、0.95，1，1.1，1.25。

Case6: $\lambda$  為-0.35； $\sigma_s$  為 0.5； $\frac{AUC_T}{AUC_R}$  比值為 0.8、0.95，1，1.1，1.25。

在此六種模式下分別模擬1000次，比較兩者在進行生物相等性檢定的型 I 誤差率及檢定力的表現。

本模擬中，在顯著水準 $\alpha=0.05$  下，型I誤差率估計的標準誤差為

$0.007 \left( = \sqrt{\frac{0.05 \times 0.95}{1000}} \right)$ ，在此超出2倍標準誤差0.037和0.064、代表顯著

水準偏離  $\alpha=0.05$ 。

## 4.2 模擬結果

由表二中可以發現參數值 $\lambda = -0.15$ ， $\sigma_s = 0.3$  的設定下，樣本數為16時，兩者的型 I 誤差率皆有控制在正常範圍內，在檢定力方面也比對數常態分布還高。樣本數為24時，對數常態分布的型 I 誤差率都已經超出2倍標準誤差為顯著偏離  $\alpha=0.05$ 。在 $\lambda = -0.35$ ， $\sigma_s = 0.3$ 的設定下，無論是樣本數為16或24時對數常態分布的型 I 誤差率都無法控制在正常範圍內，並且當隨機效應項 $S_{ik}$ 的變異數增加

$\sigma_s = 0.5$ 時，可以發現到，型 I 誤差率影更加保守的現象，而逆伽瑪分布無論是在何種設定下，型 I 誤差率的控制都在正常的範圍且檢定力的表現，都比對數常態分布更好。

## 第五章 實例分析

本文的資料根據李念純(2011)適合度檢定，針對廣義伽瑪分布所包含的分布進行適合度檢定，其結果(表五)為AUC資料為學名藥時，對數常態分布估計的p值，在顯著水準 $\alpha=0.05$ 下是在邊界，免強通過檢定，而逆伽瑪分布估計的p值是所有檢定的分布中最高的，因此本文的研究採用逆伽瑪分布來配適。以下舉一實例說明本文所提出的模式來進行生物相等性檢定。

### 5.1 資料分析

某藥廠製造一種口服降尿酸的受試藥物，為驗證其為合格的學名藥，將其與專利權已過的原廠藥，採兩期雙序列交叉設計實驗進行生體相等性評估。兩種藥物含有苯溴香豆酮(benzbromarone)成分，劑量皆為50毫克。此研究共有16位自願健康的成人參與受試，其中包含男性6位，女性10位，其年齡介於20至34歲間，身高介於155.5至173.5公分，體重範圍42至73公斤。將16位受試者隨機平分為兩組，第一序列中的8位受試者先口服原廠藥，經一週的洗除期後再服用學名藥，第二組的8位受試者則先服用學名藥，一週後再服用原廠藥。受試者服用藥物後，分別在第0.5、1、1.5、2、2.5、3、3.5、4、5、6、8、10、12及14小時等固定時間點抽血8毫升，記錄其血液中的

苯溴香豆酮濃度。並計算每位受試者所估計的樣本AUC值。

本文所提出的兩期雙序列交叉設計模式進行 SAEM 演算法求得參數估計值，步驟如下：

E 步驟為令  $m=20$ ，生成  $s_1, s_2, \dots, s_{20}$ ，20 個隨機樣本來自 (3.8) 式，

在這隨機近似過程中，計算 Q 函數 (3.9) 式且，平滑參數為

$$\delta_k = \begin{cases} 1, & 1 \leq r \leq 20 \\ \frac{1}{r-20}, & 21 \leq r \leq 40 \end{cases}$$

其中 20 代表開始做平滑(smoothing)動作的疊代次數、40 為完成參數估計的總疊代次數。M 步驟為極大化 Q 函數獲得第  $r+1$  次參數  $\theta$  的估計量。重複 E 步驟和 M 步驟 40 次，由圖一可以看出在第 20 次疊代開始至第 40 次疊代 Q 函數值就很平穩在 60，沒有大幅的變動，這也就代表參數估計值已經收斂，並且求得參數的訊息矩陣為  $\hat{I}_k$ ，參數的變異數共變異數矩陣為  $\hat{\Sigma} = \hat{I}_k^{-1}$ 。

利用 SAEM 演算法所得相關參數估計及其標準差見表三，並且建立  $\mu_T - \mu_R$  的  $100(1-2\alpha)\%$  信賴區為  $(L,U)=(-0.084,0.054)$ ，以傳統對數常態分布所建立的信賴區間為  $(-0.136,0.106)$ ，兩者所建構的區間確實都位於  $-0.223$  至  $0.223$  範圍之內，因此宣稱此藥廠製造口服降尿酸的受試藥物，與原廠專利藥具有生體相等性為合格的學名藥。

## 第六章 結論與未來研究

因為AUC其分布傾向為右偏分布，傳統上以對數常態分布加以描述，然而右偏分布不只有對數常態分布，並且根據本文研究的資料進行適合度檢定，結果顯示對數常態分布雖然有通過檢定但其P值是在邊界處，而逆伽瑪分布的P值是當中最最高，因此更適合來描述此筆資料，因而在傳統兩期雙序列交叉設計下改變誤差項的分布。

但由於誤差項不是對數常態分布，在執行兩期雙序列交叉設計模式時比傳統方法更為複雜，因為在概似函數中的隨機效應項是無法觀測到的資訊，進而必須用MCEM演算法來求得參數估計值，但由於執行MCEM演算法時生成的隨機樣本數量多，因此相當耗費時間，所以改採用SAEM演算法。

根據本文的研究，在生成500個隨機樣本下，MCEM演算法完成一次E、M步驟需2.289分鐘，而SAEM演算法在20個完成一次E、M步驟需28.798秒，假設兩者都是需要25次疊代下，參數估計值才會收斂，則MCEM演算法需耗費57分鐘，SAEM演算法要12分鐘。如果在加上估計參數標準差，差距會更加大，並且因為模擬研究需執行1000次且有不同參數設定值，因此本文的研究採用SAEM演算法。

在兩期雙序列交叉設計模式中，本文提出非傳統使用對數常態分布，根據模擬研究顯示在進行生體相等性檢定時，在不同參數值的設

定下，無論是樣本數為16或樣本數為24，型 I 誤差率皆控制在正確的範圍內，並且檢定力都比對數常態分布表現的更好。根據模擬研究發現當真實資料不是來自對數常態分布時，使用傳統方法來進行生體相等性檢定時會造成型 I 誤差率過於保守。就政府的觀點是為消費者的福利嚴格的把關，但過度保護消費者。在檢定力方面也比較低，就藥廠觀點來是不容易上市販售，在藥效的表現確實能成為學名藥，但有比較低的機率能通過學名藥的檢驗。綜合以上都不是一個好的決策法則，因此當真實資料是來自逆伽瑪分布時，以本文所提出的模式來進行生體相等性檢定會更合適。

本文因為取得的資料經檢定比較適合為逆伽瑪分布所以用它來描述AUC的分布，但所有的右偏分布，包括韋伯分布、伽瑪分布…等等皆可行，因此未來可就兩期雙序列交叉模式下考慮一個更廣的分布族來配適，如廣義伽瑪分布，其為一個廣泛的右偏分布族，如此也能得到更具穩健性的生體相等性檢定。

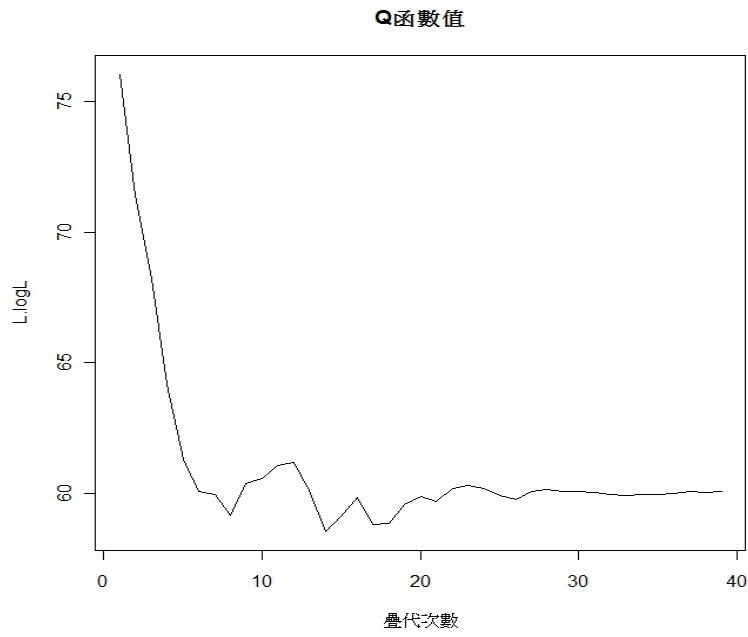
## 參考文獻

1. Berger, R.L. and Hsu, J. (1996). Bioequivalence trials, intersection-union tests and equivalence confidence sets. *Statistical Science*, 11, 283-319.
2. Cox, C. , Chu, H. , Schneider, F. and Munoz, A. (2007). “Parametric survival analysis and taxonomy of hazard functions for the generalized gamma distribution.” *Statist. in Medicine*, **26**, 4352-4374.
3. Cristian Meza · Felipe Osorio · Rolando De la Cruz (2010). Estimation in nonlinear mixed-effects models using heavy-tailed distributions. *Springer Science+Business Media, LLC* 2010
4. Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B* 39, 1-22.
5. Delyon, B., Lavielle, M. and Moulines, E. (1999) Convergence of a stochastic approximation version of the EM algorithm, *Annals of Statistics*, 27, 94-128.
6. FDA. Guidance for Industry: Statistical Approaches to Establishing Bioequivalence, Center for Drug Evaluation and Research, Food and Drug Administration, U.S. Department of Health and Human Services, 2001.

7. Louis, T.A. (1982): Finding the observed information matrix when using the EM algorithm. *J. R. Stat. Soc. Ser. B* **44**, 226–233
8. Meza, C., Jaffrézic, F., Foulley, J.-L. (2009): Estimation in the probit normal model for binary outcomes using the SAEM algorithm. *Comput. Stat. Data Anal.* **53**, 1350–1360
9. Wei, G., Tanner, M. (1990): A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *J. Am. Stat. Assoc.* **85**, 699–704
10. Yeh, K. C., and Kwan, K. C. (1978): A comparison of numerical algorithms by trapezoidal, LaGrange, and spline approximations. *J. Pharmacokinet. Biopharm.*, 6: 79-98, 1978.
11. Draper, N.R. and Smith, H. (1981): *Applied Regression Analysis*. 2<sup>nd</sup> ed., John Wiley & Sons, New York.
12. Rowland, M. and Tozer, T.N. (1980): Clinical pharmacokinetics Concepts and Applications. *Lea & Febiger, Philadelphia, PA.*
13. 李念純，一維及二維右設限存活資料的適合度檢定，國立中央大學，碩士論文，民國100年



# 附錄



圖二 SAEM 演算法 Q 函數

表二 模擬結果

$\lambda = -0.15$ $\sigma_s = 0.3$		Ratio				
		0.8	1.25	0.95	1	1.1
N=16	Inverse Gamma	0.055	0.053	0.967	1.000	0.834
	lognormal	0.051	0.046	0.904	0.966	0.782
N=24	Inverse Gamma	0.049	0.045	0.997	1.000	0.916
	lognormal	0.035	0.036	0.954	1.000	0.885
$\lambda = -0.35$ $\sigma_s = 0.3$						
N=16	Inverse Gamma	0.049	0.053	0.351	0.391	0.266
	lognormal	0.037	0.038	0.259	0.273	0.199
N=24	Inverse Gamma	0.057	0.048	0.578	0.659	0.426
	lognormal	0.035	0.038	0.515	0.586	0.372
$\lambda = -0.35$ $\sigma_s = 0.5$						
N=16	Inverse Gamma	0.045	0.047	0.260	0.283	0.210
	lognormal	0.034	0.034	0.178	0.186	0.141
N=24	Inverse Gamma	0.050	0.046	0.480	0.569	0.377
	lognormal	0.032	0.037	0.424	0.481	0.323

表三 參數估計值及標準差(逆伽瑪分布)

參數	$\mu_R$	$\mu_T$	$p$	$\lambda$	$\sigma_s$
估計值	2.481	2.466	-0.016	-0.119	0.306
估計標準差	0.047	0.048	0.021	0.020	0.059

表四 參數估計值及標準差(對數常態分布)

參數	$\mu_R$	$\mu_T$	$p$	$\lambda$	$\sigma_s$
估計值	2.459	2.474	-0.032	0.127	0.320
估計標準差	0.090	0.090	0.045	0.006	0.042

表五 適合度檢定

原廠藥				
分布	Weibull	Gamma	Lognormal	Inverse Gamma
P 值	0.031	0.057	0.149	0.350
學名藥				
分布	Weibull	Gamma	Lognormal	Inverse Gamma
P 值	0.005	0.023	0.055	0.166