Laird (1978)提出的無母數最大概似（nonparametric maximum likelihood，以下簡稱NPML）便屬於NPEB法則。

除了上述估計方法外，本文根據Escobar (1994)估計常態均數的方法，引進Dirichlet過程當G的先驗分配，提供一個新的NPEB估計法(DP)。由模擬研究得知，當均數$\lambda_1, \ldots, \lambda_n$的分配$G$符合伽碼分配的型式，利用PEB估計的$\hat{G}$當先驗分配估計$\lambda_i$有很好的效果。然而當真正$G$的型式爲少數幾個值的離散分配，NPML的$\lambda_i$估計效果顯著優於PEB。而採用DP估計$\lambda_i$的好處是，不論$G$爲何種型態，其效果總是介於NPMLE及PEB之間，並趨向其中較好的估計法。

本文第二節，介紹使用Dirichlet過程當成G先驗分配的觀念，討論在固定Dirichlet過程參數(α_0, G_0)之下，結合觀察值$Y_1, \ldots, Y_n$，估計卜瓦松均數的理論和演算法；第三節進一步討論Dirichlet過程參數(α_0, G_0)所代表的意義以及對估計的影響，並給此二參數先驗分配，改進第二節的演算法；第四節介紹三個以往估計卜瓦松均數的方法：(1) Robbins，(2) NPML，(3) PEB；第五節將DP估計法和第四節提到的三種方法作模擬比較並對Archibald (1948)提供的海石花(Armeria maritima)樣區資料作分配估計；第六節做結論。

以下將本文所使用的特殊符號作一說明，$\lambda \sim G$代表λ服從G的分配，即G爲λ的分配函數(distribution function)，$dG(\lambda) = g(\lambda)$表示G的密度函數(density function)；粗體$\boldsymbol{\lambda}$及$\boldsymbol{Y}$分別表示向量$(\lambda_1, \ldots, \lambda_n)$和$(Y_1, \ldots, Y_n)$，$\boldsymbol{\lambda}^{-i}$則表示將$\boldsymbol{\lambda}$的第$i$個元素移除。

2 Dirichlet過程

Dirichlet過程(Dirichlet Process)是一個在（機率）分配函數空間上的分配，若我們將分配函數空間上的任一種分配視爲一個「值」，針對所有「值」給定的離散分配，便稱爲「Dirichlet過程」，因此分配由Ferguson (1973)首先提出，又名Ferguson

分配。

$G \sim DP(\alpha_0, G_0)$ 表示 G 服從參數爲 α_0 和 G_0 的 Dirichlet Process 分配，可視爲分配空間中各種分配的混和型，且混和的方式由一個測度來決定，其中，G_0 表示 G 最接近的分配型式；α_0 表示我們相信分配 G 就是 G_0 的程度。G_0 和 α_0 的概念和期望值、變異數，有異曲同工之妙，想像在一個箱子裡，有著各式各樣不公正的骰子，每一種骰子視爲一種分配。某些種多、某些種少，而 Dirichlet Process 就是在各類骰子上的個數分配，G_0 指最具代表性的骰子，而 α_0 所有骰子和代表骰子的差異度。

在實際估計 λ_i 時，我們利用 Blackwell 和 MacQueen (1973) 提出的 Polya urn 結構抽樣方法來表示服從 Dirichlet 過程 G 的機率函數。若 $G \sim DP(\alpha_0, G_0)$，在 (α_0, G_0) 給定之下（G_0 爲一可抽樣的分配），Polya urn 從 G 抽出樣本 $\lambda_1, \ldots, \lambda_n$ 的方法爲

$$
\begin{aligned}
&\lambda_1 \sim G_0\text{，}\\
&\lambda_2 \mid \lambda_1 \begin{cases} \sim G_0\text{，} & \text{機率爲} \quad \dfrac{\alpha_0}{\alpha_0+1}\text{，}\\ \lambda_1\text{，} & \text{機率爲} \quad \dfrac{1}{\alpha_0+1}\text{，}\end{cases}\\
&\qquad\vdots\\
&\lambda_n \mid \lambda_1, \ldots, \lambda_{n-1} \begin{cases} \sim G_0\text{，} & \text{機率爲} \quad \dfrac{\alpha_0}{\alpha_0+n-1}\text{，}\\ \lambda_j\text{，} & \text{機率爲} \quad \dfrac{1}{\alpha_0+n-1}\text{，} \quad j = 1, \ldots, n-1\text{。}\end{cases}
\end{aligned}
$$

我們將 $\lambda_1, \ldots, \lambda_n$ 想成 n 位客人參加餐宴，依續就坐餐桌的情形，且某 i 位客人坐定後，我們將 λ_i 的值想成等同於此餐桌之桌號。第 1 位客人任意挑選一餐桌就坐 $(\lambda_1 \sim G_0)$；第 2 位客人可選擇另一新的桌子入座 $(\lambda_2 \sim G_0)$ 或和第 1 位客人坐同桌 $(\lambda_2 = \lambda_1)$，其機率分別爲 $\alpha_0/(\alpha_0+1)$ 及 $1/(\alpha_0+1)$；以此類推，第 n 位客人有機率 $\alpha_0/(\alpha_0+n-1)$ 選擇坐新桌 $(\lambda_n \sim G_0)$，有機率 $1/(\alpha_0+n-1)$ 和前面的某一位客人做同一桌 $(\lambda_n = \lambda_j,\ j = 1\ldots, n-1)$。將同一桌的人視爲同一群，直到 n 位客人都入座後，可將這 n 位客人 $\lambda_1, \ldots, \lambda_n$ 分成 k 群 $(k \leq n)$，每一群的 λ 值都相同。

在此抽樣結構下，$\lambda_1,\ldots,\lambda_n$ 的聯合機率函數便可寫成

$$dF[\lambda_1,\ldots,\lambda_n|(\alpha_0,G_0)]=\prod_{i=1}^{n}\frac{\alpha_0\cdot dG_0(\lambda_i)+\sum_{j=1}^{i-1}\delta(\lambda_i,\lambda_j)}{\alpha_0+i-1}\text{，}\tag{2.1}$$

其中，當 $\lambda_i=\lambda_j$，$\delta(\lambda_i,\lambda_j)=1$；當 $\lambda_i\neq\lambda_j$，$\delta(\lambda_i,\lambda_j)=0$。

根據 Blackwell 和 MacQueen (1973) 的證明，(2.1) 式即爲 $G\sim DP(\alpha_0,G_0)$ 之下，$\lambda_1,\ldots,\lambda_n$ 的聯合先驗分配。

Antoniak (1974) 證明若我們使用 Dirichlet 過程當成 $\lambda_1,\ldots,\lambda_n$ 的先驗分配（$G\sim DP(\alpha_0,G_0)$），則其後驗分配爲混合 Dirichlet 過程的型式。根據 Polya urn 的結構，$\lambda_1,\ldots,\lambda_n$ 的後驗分配可表示成

$$dF[\lambda_1,\ldots,\lambda_n|\boldsymbol{Y},(\alpha_0,G_0)]\propto\prod_{i=1}^{n}l(\lambda_i|Y_i)\frac{\alpha_0\cdot dG_0(\lambda_i)+\sum_{j=1}^{i-1}\delta(\lambda_i,\lambda_j)}{\alpha_0+i-1}\text{，}\tag{2.2}$$

其中 $l(\lambda_i|Y_i)$ 爲 λ_i 的概似函數。

若我們想利用 (2.2) 式估計某一個 λ_i，必須對其餘 λ（即 $\boldsymbol{\lambda}^{-i}$）積分，以得到 λ_i 的後驗分配。然而在這 n 個混合分配相乘的型式下，積分是相當困難的。爲了使估計的工作能繼續進行，Escobar (1988) 證明了另一型式的條件分配

$$dF[\lambda_i|\boldsymbol{\lambda}^{-i},\boldsymbol{Y},(\alpha_0,G_0)]\propto l(\lambda_i|Y_i)\alpha_0 dG_0(\lambda_i)+\sum_{j=1,j\neq i}^{n}l(\lambda_j|Y_i)\delta(\lambda_i,\lambda_j)\text{，}\tag{2.3}$$

如此，在給定 $\boldsymbol{Y}$ 和 $\boldsymbol{\lambda}^{-i}$ 下，λ_i 條件分配不像 (2.2) 相乘的複雜型式，取而代之的是 1 個連續分配及 $n-1$ 個退化點分配的混合型。爲了抽樣方便，我們將 (2.3) 改寫成以下的型式：

$$\lambda_i|\ \boldsymbol{\lambda}^{-i},\boldsymbol{Y}\begin{cases}\sim h(\lambda_i|Y_i)\text{，} & \text{機率正比於 } \alpha_0\cdot m(Y_i)\text{，}\\ =\lambda_j\text{，} & \text{機率正比於 } l(\lambda_j|Y_i)\text{，}\quad j\neq i\text{，}\end{cases}\qquad i=1,\ldots,n\text{，}\tag{2.4}$$

其中，$h(\lambda_i|Y_i)\propto l(\lambda_i|Y_i)dG_0(\lambda_i)$，爲 G_0 當先驗分配時 λ_i 的後驗分配；$m(Y_i)=\int h(\lambda_i|Y_i)d\lambda_i$，爲 $h(\lambda_i|Y_i)$ 的標準化常數。此外，提醒讀者，(2.4) 式中 λ_i 的條件分配僅和 $\boldsymbol{Y}$ 中的 Y_i 有關，分配 $\lambda_i|\ \boldsymbol{\lambda}^{-i},\boldsymbol{Y}$ 又可寫成 $\lambda_i|\ \boldsymbol{\lambda}^{-i},Y_i$。

有了分配(2.4)，給定其餘$\boldsymbol{\lambda}^{-i}$之下$\lambda_i$的後驗條件分配，我們利用吉氏取樣(Gibbs sampler)的過程，幫助我們間接的得到後驗邊際分配$\lambda_i|\boldsymbol{Y}$。

吉氏取樣是一種利用條件分配間接得到邊際分配的方法，它可避免掉算邊際分配時所遇到的複雜積分問題。以三個變數$\theta_1, \theta_2, \theta_3$的後驗分配爲例，假定條件分配$f(\theta_1 \mid \theta_2, \theta_3, \boldsymbol{Y})$、$f(\theta_2 \mid \theta_1, \theta_3, \boldsymbol{Y})$和$f(\theta_3 \mid \theta_1, \theta_2, \boldsymbol{Y})$爲已知可抽樣的分配，$\boldsymbol{Y} = (Y_1, Y_2, Y_3)$爲已知觀察值，吉氏取樣交替地從這三個分配抽樣，形成一個由隨機變數所組成的「吉氏數列（Gibbs sequence）」，

$$\theta_2^{(0)},\ \theta_3^{(0)},\ \theta_1^{(1)},\ \theta_2^{(1)},\ \theta_3^{(1)},\ \theta_1^{(2)},\ \theta_2^{(2)},\ \theta_3^{(2)}, \ldots, \theta_1^{(m)},\ \theta_2^{(m)},\ \theta_3^{(m)}，$$

其中除了$\theta_2^{(0)}, \theta_3^{(0)}$是隨意給定的初始值外，其餘皆以下列規則產生

$$\begin{aligned}
\theta_1^{(j+1)} &\sim f(\theta_1 \mid \theta_2^{(j)}, \theta_3^{(j)}, \boldsymbol{Y})，\\
\theta_2^{(j+1)} &\sim f(\theta_2 \mid \theta_3^{(j)}, \theta_1^{(j+1)}, \boldsymbol{Y})，\\
\theta_3^{(j+1)} &\sim f(\theta_3 \mid \theta_1^{(j+1)}, \theta_2^{(j+1)}, \boldsymbol{Y})。
\end{aligned}$$

Geman and Geman (1984)指出，在適當的條件下，當$m \to \infty$，$\theta_1^{(m)}$的分配會收斂到θ_1真正的後驗邊際分配$f(\theta_1 \mid \boldsymbol{Y})$。也就是說在$m$夠大的情況下，吉氏取樣法可以生成一個服從$f(\theta_1 \mid \boldsymbol{Y})$分配的樣本$\theta_1^{(m)}$，當然，我們也同時得到服從$f(\theta_2 \mid \boldsymbol{Y})$和$f(\theta_3 \mid \boldsymbol{Y})$分配的樣本$\theta_2^{(m)}$及$\theta_3^{(m)}$。

重複吉氏取樣L次後，可以得到L組的$(\theta_1^{(m)}, \theta_2^{(m)}, \theta_3^{(m)})$，記爲$\boldsymbol{\theta}_{(l)} = (\theta_{1(l)}, \theta_{2(l)}, \theta_{3(l)})$，$l = 1, \ldots, L$。Gelfand和Smith (1990)進一步利用這L組$\boldsymbol{\theta}_{(l)}$和已知的後驗條件分配，來估計$\theta_1$的後驗邊際分配$f(\theta_1 \mid \boldsymbol{Y})$，他們稱此法爲Rao-Blackwellization，即

$$\hat{f}(\theta_1 \mid \boldsymbol{Y}) = \frac{1}{L}\sum_{l=1}^{L} f(\theta_1 \mid \theta_{2(l)}, \theta_{3(l)}, \boldsymbol{Y})，$$

以相同的方式，我們也可以估算出$\hat{f}(\theta_2 \mid \boldsymbol{Y})$和$\hat{f}(\theta_3 \mid \boldsymbol{Y})$。

瞭解了吉氏取樣法後，以下把利用後驗條件分配(2.4)估計λ_i後驗邊際分配的步驟，做一整理。

步驟1.　設定初始値$\boldsymbol{\lambda}^{(0)} = (\lambda_1^{(0)}, \ldots, \lambda_n^{(0)})$爲$(Y_1, \ldots, Y_n)$。

步驟2.　將初始値帶入(2.4)式的條件期望値取樣。

從分配　$\lambda_1 \mid \quad \lambda_2 = \lambda_2^{(0)}, \lambda_3 = \lambda_3^{(0)}, \ldots, \lambda_n = \lambda_n^{(0)}, \boldsymbol{Y}$　抽取 $\lambda_1^{(1)}$，

從分配　$\lambda_2 \mid \lambda_1 = \lambda_1^{(1)}, \quad \lambda_3 = \lambda_3^{(0)}, \ldots, \lambda_n = \lambda_n^{(0)}, \boldsymbol{Y}$　抽取 $\lambda_2^{(1)}$，

從分配　$\lambda_3 \mid \lambda_1 = \lambda_1^{(1)}, \lambda_2 = \lambda_2^{(1)}, \quad \ldots, \lambda_n = \lambda_n^{(0)}, \boldsymbol{Y}$　抽取 $\lambda_3^{(1)}$，

$\vdots$

從分配　$\lambda_n \mid \lambda_1 = \lambda_1^{(1)}, \lambda_2 = \lambda_2^{(1)}, \ldots, \lambda_{n-1} = \lambda_{n-1}^{(1)}, \quad \boldsymbol{Y}$　抽取 $\lambda_n^{(1)}$。

經過了這n步抽樣，我們稱完成一次疊代，得到$\boldsymbol{\lambda}^{(1)} = (\lambda_1^{(1)}, \ldots, \lambda_n^{(1)})$，再將其入(2.4)式取樣，可得$\boldsymbol{\lambda}^{(2)} = (\lambda_1^{(2)}, \ldots, \lambda_n^{(2)})$。一直重複此疊代動作$m$次後，我們得到$\boldsymbol{\lambda}^{(m)} = (\lambda_1^{(m)}, \ldots, \lambda_n^{(m)})$。

步驟3.　步驟2重複L次，我們得到L組$\boldsymbol{\lambda}^{(m)} = (\lambda_1^{(m)}, \ldots, \lambda_n^{(m)})$，記爲$\boldsymbol{\lambda}_{(l)} = (\lambda_{1(l)}, \ldots, \lambda_{n(l)})$，$l = 1, \ldots, L$。

於是λ_i的後驗分配估計如下：

$$\hat{f}(\lambda_i \mid \boldsymbol{Y}) = \frac{1}{L}\sum_{l=1}^{L} f(\lambda_i \mid \boldsymbol{\lambda}_{(l)}^{-i}, \boldsymbol{Y}), \qquad i = 1, \ldots, n, \tag{2.5}$$

因爲(2.4)式的λ_i後驗條件分配僅和$\boldsymbol{Y}$中的Y_i有關，我們可將(2.5)式改寫成

$$\hat{f}(\lambda_i \mid Y_i) = \frac{1}{L}\sum_{l=1}^{L} f(\lambda_i \mid \boldsymbol{\lambda}_{(l)}^{-i}, Y_i), \qquad i = 1, \ldots, n, \tag{2.6}$$

在貝氏損失函數$L(\hat{\lambda}_i, \lambda_i) = (\hat{\lambda}_i - \lambda_i)^2$之下，貝氏估計量爲後驗分配的期望値，因此分配(2.6)的期望値，爲使用Dirichlet過程當先驗分配，λ_i的估計量，即

$$\hat{\lambda}_i = \hat{E}\,[\lambda_i \mid Y_i] = \frac{1}{L}\sum_{l=1}^{L} E[\lambda_i \mid \boldsymbol{\lambda}_{(l)}^{-i}, Y_i], \qquad i = 1, \ldots, n, \tag{2.7}$$

其中，$E[\lambda_{i(l)} \mid \boldsymbol{\lambda}_{(l)}^{(-i)}, Y_i]$爲條件分配(2.4)的期望値。