有了λ_i的估計先驗分配$\Gamma(\hat{a},\hat{b})$，接著再利用貝氏定理，我們得到λ_i的後驗分配

$$\begin{aligned} h(\lambda_i \mid Y_i) &\propto l(\lambda_i \mid Y_i) \cdot g(\lambda_i) \\ &\propto e^{-\lambda_i} \frac{\lambda_i^{Y_i}}{Y_i!} \cdot \frac{\lambda_i^{\hat{a}-1} e^{-\lambda_i/\hat{b}}}{\Gamma(\hat{a})\hat{b}^{\hat{a}}} \\ &\propto \lambda_i^{(\hat{a}+Y_i)-1} e^{-\lambda_i/(\frac{\hat{b}}{1+\hat{b}})} \\ &\equiv \Gamma\left(\hat{a}+Y_i, \tfrac{\hat{b}}{1+\hat{b}}\right) \text{。} \end{aligned} \tag{4.7}$$

同樣地，我們求(4.7)式的期望值，並將之當成λ_i的貝氏估計量

$$\hat{\lambda}_i^{PEB} = E(\lambda_i \mid Y_i) = \tfrac{(\hat{a}+Y_i)\hat{b}}{1+\hat{b}} = \tfrac{\hat{a}\hat{b}}{1+\hat{b}} + \tfrac{Y_i\hat{b}}{1+\hat{b}} = \tfrac{1}{1+\hat{b}}\overline{Y} + \tfrac{\hat{b}}{1+\hat{b}}Y_i \text{。} \tag{4.8}$$

由(4.8)式可知，λ_i的PEB估計量爲所有觀察值平均$\overline{Y}$和Y_i加權平均的結果，其權重分別爲$1/(1+\hat{b})$、$\hat{b}/(1+\hat{b})$，且和爲1。當資料$\boldsymbol{Y}$很集中時，變異數S^2小，$\hat{b}$跟著小，$1/(1+\hat{b})$相對較大，此時用較多其他觀察值的資訊來幫助估計λ_i；當資料$\boldsymbol{Y}$較分散時，$1/(1+\hat{b})$相對較小，此時相信λ_i應該和Y_i以外的觀察值關係不大。

5 模擬研究和實例探討

5.1 模擬研究

從已知分配G中抽取λ_i，再以此λ_i當成卜瓦松的參數，抽取Y_i值，其中$i=1,\ldots,n$。爲了不讓估計方法受限於特定的觀察值，我們取$n=15$、30、50三種樣本大小，且分別討論G爲「伽瑪分配」及「特定伯努力分配」的狀況，型式如下：

表2. G為伽瑪分配以及特定伯努力分配的機率密度函數

G爲伽瑪分配	G爲特定伯努力分配
$f(\lambda_i) = \dfrac{\lambda_i^{1-1} e^{-\lambda_i/b}}{\Gamma(1) b^1}$，	$P(\lambda_i) = \begin{cases} 0.8\text{，} & \lambda_i = 0.5\text{，} \\ 0.2\text{，} & \lambda_i = c\text{，} \\ 0\text{，} & \text{otherwise}\text{，} \end{cases}$
其中，$b=1$、2、3.16。	其中，$c=1.5$、5.5、10.5。

在G爲伽瑪分配$\Gamma(1,b)$下，將b分爲1、2、3.16三種情形，探討λ_i變異數大小對估計的影響；在G爲特定伯努力分配下，取c爲1.5、5.5、10.5三種情況，討論兩群λ_i分離度高低對估計的影響。混合不同大小和不同分配，觀察值便有18種不同型態。

實際模擬操作，當使用NPML估計法時，我們取疊代次數$m = 500$；在使用DP估計法時，我們重複$100(= L)$組吉氏法疊代，每組疊代$16(= m)$次，而在估算(3.5)式(α_0, G_0)對觀察值$\boldsymbol{Y}$的概似函數時，抽取$1024(= K)$組$\boldsymbol{\lambda}$來平均。

由於模擬能掌握$\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)$真實的值，只要拿各種方法的估計值和真實值做比較，便可探討這些估計法的好壞。若估計式爲$\hat{\boldsymbol{\lambda}} = (\hat{\lambda_1}, \ldots, \hat{\lambda_n})$，我們利用均方誤(mean square error)當成比較的準則，即$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{\lambda_i} - \lambda_i)^2$。此外，爲了瞭解各種估計方法的穩定性，對於相同型態的觀察值，重複取樣50次，計算每一次的均方誤，記爲$MSE_{(1)}, \ldots, MSE_{(50)}$。下表列出的平均和變異數，即爲$MSE_{(1)}, \ldots, MSE_{(50)}$的平均和變異數。

表3. 卜瓦松均數$\boldsymbol{\lambda}$符合表2伽碼分配下，各種估計法之

50次均方誤(MSE)的平均及變異數

$\underline{n = 15}$

估計法	$b = 1$		$b = 2$		$b = 3.16$	
	平均	變異數	平均	變異數	平均	變異數
Y_i	0.98	0.35	2.33	3.52	3.72	4.04
Robbins	2.11	1.66	9.29	34.67	18.43	83.70
NPML	0.71	0.15	2.35	4.08	3.37	3.61
PEB	0.60	0.10	1.89	1.81	2.82	1.89
DP	0.79	0.39	2.33	3.55	3.64	3.82

$\underline{n = 30}$

估計法	$b = 1$		$b = 2$		$b = 3.16$	
	平均	變異數	平均	變異數	平均	變異數
Y_i	0.99	0.21	1.99	0.67	3.17	2.19
Robbins	2.08	1.38	7.43	6.24	20.84	64.08
NPML	0.66	0.13	1.68	0.69	3.47	2.86
PEB	0.54	0.05	1.49	0.54	2.73	1.89
DP	0.66	0.12	1.63	0.75	3.13	2.26

$\underline{n = 50}$

估計法	$b = 1$		$b = 2$		$b = 3.16$	
	平均	變異數	平均	變異數	平均	變異數
Y_i	1.08	0.15	1.85	0.24	3.00	1.18
Robbins	2.06	0.53	8.00	9.11	19.47	33.10
NPML	0.68	0.09	1.58	0.28	2.84	0.90
PEB	0.57	0.05	1.37	0.20	2.41	0.68
DP	0.60	0.13	1.33	0.17	2.67	0.82

表4. 卜瓦松均數$\boldsymbol{\lambda}$符合表2特定伯努力分配分配下，各種估計法之50次均方誤(MSE)的平均及變異數

	$n=15$					
估計法	$c=1.5$		$c=5.5$		$c=10.5$	
	平均	變異數	平均	變異數	平均	變異數
Y_i	0.74	0.11	1.76	1.06	2.34	2.86
Robbins	0.68	0.04	6.21	10.53	20.49	93.42
NPML	0.27	0.06	1.38	1.54	1.18	2.03
PEB	0.27	0.04	1.23	0.69	2.18	2.28
DP	0.42	0.09	1.48	1.39	1.23	1.81

	$n=30$					
估計法	$c=1.5$		$c=5.5$		$c=10.5$	
	平均	變異數	平均	變異數	平均	變異數
Y_i	0.64	0.07	1.54	0.31	2.44	1.39
Robbins	0.65	0.02	6.59	3.86	22.47	84.73
NPML	0.21	0.01	0.90	0.30	0.86	0.72
PEB	0.19	0.01	1.16	0.12	2.29	1.14
DP	0.20	0.02	0.96	0.24	0.89	0.59

	$n=50$					
估計法	$c=1.5$		$c=5.5$		$c=10.5$	
	平均	變異數	平均	變異數	平均	變異數
Y_i	0.75	0.09	1.60	0.35	2.54	1.93
Robbins	0.66	0.02	6.63	2.49	22.43	31.77
NPML	0.21	0.02	0.84	0.29	0.80	1.03
PEB	0.19	0.01	1.14	0.17	2.31	1.31
DP	0.22	0.03	0.91	0.25	0.80	0.86

由表3、表4可知，直接用觀察值Y_i對λ_i做估計的效果並不理想，因爲這樣的估計忽略了先驗分配G；然而當我們用Robbins估計法時，誤差更甚於直接用觀察值的估計，且在觀察值不集中的狀況下，誤差竟爲觀察值Y_i估計法的數倍。因此以下排除此二種不好的估計法，針對NPML、PEB和DP三種估計法分析討論。

當G爲伽瑪分配(表2)時，在$b = 1$（亦即觀察值$\boldsymbol{Y}$較集中）的情況下，$n(= 15)$較小時，DP估計法誤差較大(0.79)，然而一旦n增至30、50，DP估計法的表現便和NPML、PEB相近($\doteq 0.6$)。在$b = 2$、3.16（亦即觀察值$\boldsymbol{Y}$較分散）的情況下，PEB的估計表現較佳，爲所有估計法中平均誤差最小的，這結果和我們預期相符，因爲PEB估計法猜中了$\boldsymbol{\lambda}$先驗分配的型式；隨著n變大，DP估計法的的誤差表現向PEB估計法接近，甚至在$n = 50$，$b = 2$時表現（即MSE的平均）要比PEB好（$1.33 < 1.37$）；雖然NPML的估計法亦隨著n變大，估計誤差變小，但表現始終不如DP估計法。從變異數來看，在$n(= 15)$較小時，DP及NPML估計法較PEB不穩定，然而這種情況在n增加到50時，便有明顯改善。

當G爲特殊伯努力分配(表3)，同樣地，在$c = 1.5$（亦即觀察值$\boldsymbol{Y}$較集中）的情況下，n較小時，DP估計法誤差較大(0.42)，不過隨著n增加，表現會和NPML、PEB相近($\doteq 0.2$)。在$c = 5.5$、10.5（亦即觀察值$\boldsymbol{Y}$集中成分開的兩群），NPML的估計法反過來優於PEB估計法，此現象在此兩群觀察值距離大($c = 10.5$)時，非常明顯，如$n = 50$，MSE的平均分別爲0.8及2.13；隨著n變大，DP估計法的的誤差表現也反過來接近NPML估計法，在$n = 50$，$c = 10.5$的情況下，誤差甚至可以相等。此外，觀察變異數得知，當G爲特殊伯努力分配，NPML、PEB及DP的穩定性表現差不多。

綜合以上，我們得到一個興奮的結果，在觀察值$\boldsymbol{Y}$夠多時（建議≥ 50），DP估計法的估計表現，不單單只是介於PEB和NPML間，它能向最佳的估計法趨近。這性質讓我們實際在估計「未知型態」的$\boldsymbol{\lambda}$時，能夠有較準確的結果。

5.2 實例探討

將一個區域劃分成N塊面積相等的區塊，利用抽樣方法隨機抽取$n(< N)$個區塊來估計區域內所有觀察值的個數，稱爲樣區法(quadrat method)。以往應用樣區法來進行某種植物分析時，都假設植物的生長沒有地域性，隨機地散佈在要估計的區域內，且將所有樣區觀察值視爲來自同一個均數的卜瓦松分配。然而Archibald (1948)對數種植物進行卜瓦松配適分析，證明這個隨機的假設並不全然成立，解釋了大多數植物有群集的趨勢。我們針對Archibald (1948)提供的海石花(Armeria maritima)樣區資料，將每個樣區視爲具有個別均數的卜瓦松分配，利用DP法來估計個別樣區均數，進而估計區域內海石花個數之分配。

表5. 100個樣區的海石花個數紀錄資料以及每一樣區的均數估計

樣區花數	對應樣區數	樣區花數卜瓦松均數估計(DP)	估計分配的理論次數
0	57	0.12	52.04
1	6	2.56	10.95
2	12	2.98	7.97
3	5	3.38	8.71
4	5	3.76	7.40
5	5	4.12	5.43
6	7	4.46	3.51
7	1	4.83	1.99
8	—	—	1.04
9	1	5.82	0.54
10	1	6.42	0.24
≥ 11	—	—	0.18
加總	100	—	100

爲了DP法的使用，我們將原始資料（表5）整理成(1.1)卜瓦松模型的型式，令$Y_1,\ldots,Y_{57}$的值爲0，$Y_{58},\ldots,Y_{63}$的值爲1，以此類推，$Y_{99}=9$，$Y_{100}=10$；而相對應的估計值$\hat{\lambda}_1,\ldots,\hat{\lambda}_{57}$爲0.12，$\hat{\lambda}_{58},\ldots,\hat{\lambda}_{63}$爲2.56，以此類推，$\hat{\lambda}_{99}=5.82$，$\hat{\lambda}_{100}=6.42$。

有了每個樣區的估計均數，我們利用混和卜瓦松分配估計區域內海石花的個數分配，型式如下

$$\hat{f}(x)=\sum_{i=1}^{10} w_i\cdot e^{-\lambda_i^*}\frac{\lambda_i^{*x}}{x!}\,,\qquad x=0,1,2,\ldots\,, \tag{5.1}$$

其中，$\boldsymbol{w}=(0.57,\ \ 0.06,\ \ 0.12,\ \ 0.05,\ \ 0.05,\ \ 0.05,\ \ 0.07,\ \ 0.01,\ \ 0.01,\ \ 0.01)$，

$\boldsymbol{\lambda}^*=(0.12,\ \ 2.56,\ \ 2.98,\ \ 3.38,\ \ 3.76,\ \ 4.12,\ \ 4.46,\ \ 4.83,\ \ 5.82,\ \ 6.42)$。

利用(5.1)式，重複抽取1000組樣本，每組100個觀察值，次數平均結果如表5.的「估計分配的理論次數」。在虛無假設H_0表示觀察值服從(5.1)式分配下，卡方配適度檢定(goodness-of-fit test)統計量爲

$$\begin{aligned}\psi&=\left[\frac{57^2}{52.04}+\frac{6^2}{10.95}+\cdots+\frac{1^2}{0.18}\right]-100\\&=15.122<19.675=\chi^2_{0.95}(11)\,,\end{aligned} \tag{5.2}$$

若顯著水準爲95%，我們不拒絕虛無假設，從而推論區域內海石花的個數分配爲(5.1)之型式。

6 結論

1. 從模擬結果可知，不論G爲何種型態，DP估計法總是介於NPML及PEB之間，並趨向其中較好的估計法。

2. 採用Dirichlet過程當成G的先驗分配，爲一種特別的NPEB法則，不僅使$\boldsymbol{\lambda}$的先驗分配不受限於某種分配型式，當G偏向某種分配時，還可以藉由G_0的選擇，融入此資訊。