取後不放回的抽樣方式

在單群落之情況中,假設目標區域包含S種共同物種,並將該地區分為T個大致相等的區塊。若在取樣區塊中發現該物種,則被紀錄為存在,反之則為不存在,針對群落進行取後不放回之隨機抽樣,分別抽取 t 的區塊數,僅記錄每個採樣樣本中物種的發生率。令 ψ_i 與是物種 i 在樣本群落中所佔的區塊數量 (i=1,2,...,S)。在此, ψ_i 可作為衡量物種 i 在該區域中分佈程度使用。且為了確保所調查區域中的所有區塊都被調查時,所有的物種都被考慮在內,機率模型中的參數必須大於零。綜上所述,可對 ψ_i 進行假設:遵循參數T和 π_i 的零 截尾 Beta 二項分佈 (zero-truncated beta-binomial distribution) (Shen and He, 2008):

$$P(\psi_i \mid \pi_i) = {T \choose \psi_i} \frac{\pi_i^{\psi_i} (1 - \pi_i)^{T - \psi_i}}{1 - (1 - \pi_i)^T}, i = 1, 2, ..., S$$

其中, π_i 表示物種i的樣本中物種出現與否的檢測率。當取後不放回之隨機抽樣從T個區塊中抽出t 個區塊,並且每個樣本區塊中僅記錄物種的存在與否,以形成逐種樣本發生矩陣 X_i ,應遵循超幾何分佈 (hypergeometric distribution):

$$P(X_i = k | \psi_i) = \frac{\binom{\psi_i}{k} \binom{T - \psi_i}{t - k}}{\binom{T}{t}}, i = 1, 2, \dots, S$$

又 $P(X_i = k | \psi_i)$ 與 ψ_i 有關, X_i 來自於 π_i ,因此可推導出:

$$P(X_i = k | \pi_i) = \sum_{\psi_i} P(X_i = k | \psi_i) P(\psi_i | \pi_i)$$

$$= {t \choose k} \frac{\pi_i^k (1 - \pi_i)^{t-k}}{1 - (1 - \pi_i)^T} \frac{(1 - \pi_i)^T I(k = 0)}{1 - (1 - \pi_i)^T}$$

其中,I(A)為指標函數,表示若出現A情況時,則該式為1,反之則即為0。並假設 π ;為一來自beta分佈的隨機樣本,可將式式子表示為:

$$f(\pi_i) = G(\alpha, \beta, T)(1 - (1 - \pi_i)^T)\pi_i^{\alpha - 1}(1 - \pi_i)^{\beta - 1}, 0 < \pi_i < 1$$

其中
$$G(\alpha, \beta, T) = \left[\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} - \frac{\Gamma(\alpha)\Gamma(\beta+T)}{\Gamma(\alpha+\beta+T)}\right]^{-1}$$
。

後隨邊際分佈 X_i 可通過 $P(X_i = k | \psi_i)$ 獲得,故:

$$p_{k} = P(X_{i} = k) = \int_{0}^{1} P(X_{i} = k \mid \pi_{i}) f(\pi_{i}) d\pi_{i}$$

$$= \begin{cases} G(\alpha, \beta, T) \left[\frac{\Gamma(\alpha)\Gamma(t + \beta)}{\Gamma(t + \alpha + \beta)} - \frac{\Gamma(\alpha)\Gamma(\beta + T)}{\Gamma(\alpha + \beta + T)} \right], & \text{if } k = 0 \\ G(\alpha, \beta, T) \left(t \right) \frac{\Gamma(\alpha + k)\Gamma(t + \beta - k)}{\Gamma(t + \alpha + \beta)}, & \text{if } k > 0 \end{cases}$$

$$(4)$$

將其推廣至兩群落:

$$p_{kl} = P(X_i = k)P(Y_i = l) =$$

$$\begin{cases} G_1 \left[\frac{\Gamma(\alpha_1)\Gamma(t_1+\beta_1)}{\Gamma(t_1+\alpha_1+\beta_1)} - \frac{\Gamma(\alpha_1)\Gamma(\beta_1+T_1)}{\Gamma(\alpha_1+\beta_1+T_1)} \right] G_2 \left[\frac{\Gamma(\alpha_1)\Gamma(t_1+\beta_1)}{\Gamma(t_1+\alpha_1+\beta_1)} - \frac{\Gamma(\alpha_1)\Gamma(\beta_1+T_1)}{\Gamma(\alpha_1+\beta_1+T_1)} \right], & if \ k = 0 \ and \ l = 0 \\ G_1 \binom{t_1}{k} \frac{\Gamma(\alpha_1+k)\Gamma(t_1+\beta_1-k)}{\Gamma(t_1+\alpha_1+\beta_1)} G_2 \binom{t_2}{l} \frac{\Gamma(\alpha_2+l)\Gamma(t_2+\beta_2-l)}{\Gamma(t_2+\alpha_2+\beta_2)}, & if \ k > 0 \ and \ l > 0 \\ G_1 \left[\frac{\Gamma(\alpha_1)\Gamma(t_1+\beta_1)}{\Gamma(t_1+\alpha_1+\beta_1)} - \frac{\Gamma(\alpha_1)\Gamma(\beta_1+T_1)}{\Gamma(\alpha_1+\beta_1+T_1)} \right] G_2 \binom{t_2}{l} \frac{\Gamma(\alpha_2+l)\Gamma(t_2+\beta_2-l)}{\Gamma(t_2+\alpha_2+\beta_2)}, & if \ k = 0 \ and \ l > 0 \end{cases}$$

其中 p_{kl} 為量樣本的物種豐富度正好分別為k和l的平均機率。令 Q_{kl} = $\sum_{i=1}^{S} I(X_i = k \text{ and } Y_i = l) \text{ 式在樣本中第一群落出現} k 次且第二群即出現<math>l$ 次的區塊數,則 D_{12} 為樣本中觀測到的共同物種數量, $D_{12} = \sum_{k=1}^{T_1} \sum_{l=1}^{T_2} Q_{kl}$ 。

Chiu (2022) 基於 Good-Turing 頻率公式與柯西不等式 (Cauchy-Schwarz inequality) 之概念,針對單一群落的估計得出近似式: $\frac{E(Q_0)}{E(Q_1)} \ge \frac{E(Q_1)}{2E(Q_2)} \ge$

 $\frac{E(Q_2)}{3E(Q_3)}$...,其中 Q_k 為出現k個區塊的物種數。從中可以得知,在物種估計時,採取出現較少次的物種,可以更多提供未出現物種的資訊,有助於縮小物種豐富度的估計結果。根據等式 (4),給出未觀測到的豐富度的期望值、唯一值和重複值:

$$E(Q_0) = P(X_i = 0) = S_{12} \times G(\alpha, \beta, T) \left[\frac{\Gamma(\alpha)\Gamma(t+\beta)}{\Gamma(t+\alpha+\beta)} - \frac{\Gamma(\alpha)\Gamma(\beta+T)}{\Gamma(\alpha+\beta+T)} \right]$$

$$E(Q_1) = P(X_i = 1) = S_{12} \times G(\alpha, \beta, T) {t \choose 1} \frac{\Gamma(\alpha+1)\Gamma(t+\beta-1)}{\Gamma(t+\alpha+\beta)}$$

$$E(Q_2) = P(X_i = 2) = S_{12} \times G(\alpha, \beta, T) {t \choose 2} \frac{\Gamma(\alpha+2)\Gamma(t+\beta-2)}{\Gamma(t+\alpha+\beta)}$$

並將其推廣至兩群落:

$$\begin{split} E(Q_{0+}) &= S_{12} \times P(X_i = 0) \times \left(1 - P(Y_i = 0)\right) \\ &= S_{12} \times G(\alpha_1, \beta_1, T_1) \\ &\times \left[\frac{\Gamma(\alpha_1)\Gamma(t_1 + \beta_1)}{\Gamma(t_1 + \alpha_1 + \beta_1)} - \frac{\Gamma(\alpha_1)\Gamma(\beta_1 + T_1)}{\Gamma(\alpha_1 + \beta_1 + T_1)} \right] \\ &\times \left\{ 1 - G(\alpha_2, \beta_2, T_2) \left[\frac{\Gamma(\alpha_2)\Gamma(t_2 + \beta_2)}{\Gamma(t_2 + \alpha_2 + \beta_2)} - \frac{\Gamma(\alpha_2)\Gamma(\beta_2 + T_2)}{\Gamma(\alpha_2 + \beta_2 + T_2)} \right] \right\} \\ E(Q_{1+}) &= S_{12} \times P(X_i = 1) \times \left(1 - P(Y_i = 0)\right) \\ &= S_{12} \times G(\alpha_1, \beta_1, T_1) \left(\frac{t}{1} \right) \left(\frac{\Gamma(\alpha_1 + 1)\Gamma(t_1 + \beta_1 - 1)}{\Gamma(t_1 + \alpha + \beta_1)} \right) \\ &\times \left\{ 1 - G(\alpha_2, \beta_2, T_2) \left[\frac{\Gamma(\alpha_2)\Gamma(t_2 + \beta_2)}{\Gamma(t_2 + \alpha_2 + \beta_2)} - \frac{\Gamma(\alpha_2)\Gamma(\beta_2 + T_2)}{\Gamma(\alpha_2 + \beta_2 + T_2)} \right] \right\} \\ E(Q_{2+}) &= S_{12} \times P(X_i = 2) \times \left(1 - P(Y_i = 0)\right) \\ &= S_{12} \times G(\alpha_1, \beta_1, T_1) \times \left(\frac{t}{2} \right) \left(\frac{\Gamma(\alpha_1 + 2)\Gamma(t_1 + \beta_1 - 2)}{\Gamma(t_1 + \alpha + \beta_1)} \right) \\ &\times \left\{ 1 - G(\alpha_2, \beta_2, T_2) \left[\frac{\Gamma(\alpha_2)\Gamma(t_2 + \beta_2)}{\Gamma(t_2 + \alpha_2 + \beta_2)} - \frac{\Gamma(\alpha_2)\Gamma(\beta_2 + T_2)}{\Gamma(\alpha_2 + \beta_2 + T_2)} \right] \right\} \end{split}$$

將 α 設定為 1 ,且 $T \gg \beta$ 。成立以下近似值:

$$\frac{E(Q_{0+})}{E(Q_{1+})} = \left(\frac{\beta_1 + t_1 - 1}{t_1} \frac{T_1 - t_1}{T_1 + \beta_1}\right) \approx \left[\left(\frac{\beta_1}{t_1} + 1\right) \frac{(T_1 - t_1)}{T_1 + \beta_1}\right]$$
$$\frac{E(Q_{1+})}{E(Q_{2+})} = \frac{t_1 + \beta_1 - 2}{t_1 - 1} \approx \left(\frac{\beta_1}{t_1} + 1\right)$$

得
$$\beta_1 = \left(\frac{E(Q_{1+}) - E(Q_{2+})}{E(Q_{2+})}\right) t_1$$
,代入 $E(Q_{0+}) = \frac{E(Q_{1+})^2}{E(Q_{2+})} \left(\frac{(T_1 - t_1)}{T_1 + \beta_1}\right)$ 。

得:

$$E(Q_{0+}) = \begin{cases} \frac{t_1 - 1}{t_1} \frac{(1 - q)E(Q_{1+})}{E(Q_{2+}) + q\big(E(Q_{1+}) - E(Q_{2+})\big)^+}, & if \ E(Q_{2+}) > 0 \\ \frac{t_1 - 1}{t_1} \frac{1 - q}{q} (E(Q_{1+}) - 1), & if \ E(Q_{2+}) = 0 \end{cases}$$

 $(A)^{+}$ 表示:若A > 0時,則等於 $(A)^{+} = A$;若 $A \leq 0$,則等於 $(A)^{+} = 1$ 。

同理
$$E(Q_{0+})$$
 也依此證明,得 $\beta_2 = \left(\frac{E(Q_{+1}) - E(Q_{+2})}{E(Q_{+2})}\right) t_2$:

$$E(Q_{+0}) = \begin{cases} \frac{t_2 - 1}{t_2} \frac{(1 - q)E(Q_{+1})}{E(Q_{+2}) + q(E(Q_{+1}) - E(Q_{+2}))^+}, & \text{if } E(Q_{+2}) > 0 \\ \frac{t_2 - 1}{t_2} \frac{1 - q}{q}(E(Q_{+1}) - 1), & \text{if } E(Q_{+2}) = 0 \end{cases}$$

又:

$$\begin{split} E(Q_{00}) &= S_{12} \times P(X_i = 0) \times P(Y_i = 0) \\ &= S_{12}G(\alpha_1, \beta_1, T_1)G(\alpha_2, \beta_2, T_2) \\ &\times \left[\frac{\Gamma(\alpha_1)\Gamma(t_1 + \beta_1)}{\Gamma(t_1 + \alpha_1 + \beta_1)} - \frac{\Gamma(\alpha_1)\Gamma(\beta_1 + T_1)}{\Gamma(\alpha_1 + \beta_1 + T_1)} \right] \\ &\times \left[\frac{\Gamma(\alpha_2)\Gamma(t_2 + \beta_2)}{\Gamma(t_2 + \alpha_2 + \beta_2)} - \frac{\Gamma(\alpha_2)\Gamma(\beta_2 + T_2)}{\Gamma(\alpha_2 + \beta_2 + T_2)} \right] \\ E(Q_{11}) &= S_{12} \times P(X_i = 1) \times P(Y_i = 1) \\ &= S_{12} \times G(\alpha_1, \beta_1, T_1)G(\alpha_2, \beta_2, T_2) \\ &\times \binom{t_1}{1} \binom{t_2}{1} \left(\frac{\Gamma(\alpha_1 + 1)\Gamma(t_1 + \beta_1 - 1)}{\Gamma(t_1 + \alpha_1 + \beta_1)} \right) \left(\frac{\Gamma(\alpha_2 + 1)\Gamma(t_2 + \beta_2 - 1)}{\Gamma(t_2 + \alpha_2 + \beta_2)} \right) \\ E(Q_{22}) &= S_{12} \times P(X_i = 2) \times P(Y_i = 2) \\ &= S_{12} \times G(\alpha_1, \beta_1, T_1)G(\alpha_2, \beta_2, T_2) \\ &\times \binom{t_1}{2} \binom{t_2}{2} \left(\frac{\Gamma(\alpha_1 + 2)\Gamma(t_1 + \beta_1 - 2)}{\Gamma(t_1 + \alpha_1 + \beta_1)} \right) \left(\frac{\Gamma(\alpha_2 + 2)\Gamma(t_2 + \beta_2 - 2)}{\Gamma(t_2 + \alpha_2 + \beta_2)} \right) \end{split}$$

並成立以下近似值:

$$\begin{split} \frac{E(Q_{00})}{E(Q_{11})} &= \bigg(\frac{\beta_1 + t_1 - 1}{t_1} \frac{T_1 - t}{T_1 + \beta_1}\bigg) \bigg(\frac{\beta_2 + t_2 - 1}{t_2} \frac{T_2 - t_2}{T_2 + \beta_2}\bigg) \\ &\approx \bigg[\bigg(\frac{\beta_1}{t_1} + 1\bigg) \frac{(T_1 - t_1)}{T_1 + \beta_1}\bigg] \bigg[\bigg(\frac{\beta_2}{t_2} + 1\bigg) \frac{(T_2 - t_2)}{T_2 + \beta_2}\bigg] \end{split}$$

後,並加入 $\left(\frac{t_1-1}{t_1}\frac{t_2-1}{t_2}\right)$ 對估計式進行調整,最終得估計式:

$$S_{wBB1} = D_{12} + E(Q_{00}) + E(Q_{0+}) + E(Q_{+0})$$

其中:

$$\begin{split} E(Q_{00}) &= \left(\frac{t_1 - 1}{t_1} \frac{t_2 - 1}{t_2}\right) \times \left[\left(\frac{\beta_1}{t_1} + 1\right) \frac{(T_1 - t_1)}{T_1 + \beta_1}\right] \times \left[\left(\frac{\beta_2}{t_2} + 1\right) \frac{(T_2 - t_2)}{T_2 + \beta_2}\right] E(Q_{11}) \\ &\stackrel{\text{if }}{=} \left(\frac{\beta_1}{t_1} + \frac{(1 - q)E(Q_{2+})}{E(Q_{2+})}\right) t_1 \\ & \qquad \qquad \left\{\beta_2 &= \left(\frac{E(Q_{1+}) - E(Q_{2+})}{E(Q_{2+})}\right) t_2\right. \\ & \qquad \qquad E(Q_{0+}) &= \begin{cases} \frac{t_1 - 1}{t_1} \frac{(1 - q)E(Q_{1+})^2}{E(Q_{2+}) + q(E(Q_{1+}) - E(Q_{2+}))^+}, & \text{if } E(Q_{2+}) > 0 \\ & \qquad \qquad \frac{t_1 - 1}{t_1} \frac{1 - q}{q}(E(Q_{1+}) - 1), & \text{if } E(Q_{2+}) = 0 \end{cases} \\ & \qquad \qquad E(Q_{+0}) &= \begin{cases} \frac{t_2 - 1}{t_2} \frac{(1 - q)E(Q_{+1})^2}{E(Q_{+2}) + w_2(E(Q_{+1}) - E(Q_{+2}))^+}, & \text{if } E(Q_{+2}) > 0 \\ & \qquad \qquad \frac{t_2 - 1}{t_2} \frac{1 - q}{q}(E(Q_{+1}) - 1), & \text{if } E(Q_{+2}) = 0 \end{cases} \end{split}$$

並在 S_{wBB1} 的基礎上,加入 Q_{12} 、 Q_{21} 對 β_1 、 β_2 的估計進行修正,成立以下近似值:

$$\frac{E(Q_{11})}{E(Q_{22})} \approx \left(\frac{\beta_1}{t_1} + 1\right) \left(\frac{\beta_2}{t_2} + 1\right)$$

經由式 (12) 與 式 (17) 推得出:

$$\begin{cases} \beta_1 = \frac{E(Q_{11})}{E(Q_{22})} \left(\frac{t_2}{\beta_2 + t_2}\right) t_1 - t_1 \\ \beta_2 = \frac{E(Q_{11})}{E(Q_{22})} \left(\frac{t_1}{\beta_1 + t_1}\right) t_2 - t_2 \end{cases}$$

又,

$$\begin{split} E(Q_{12}) &= S_{12} \times P(X_i = 1) \times P(Y_i = 2) \\ &= S_{12} \times G(\alpha_1, \beta_1, T_1) G(\alpha_2, \beta_2, T_2) \\ &\times \binom{t_1}{1} \binom{t_2}{2} \left(\frac{\Gamma(\alpha_1 + 1) \Gamma(t_1 + \beta_1 - 1)}{\Gamma(t_1 + \alpha_1 + \beta_1)} \right) \left(\frac{\Gamma(\alpha_2 + 2) \Gamma(t_2 + \beta_2 - 2)}{\Gamma(t_2 + \alpha_2 + \beta_2)} \right) \end{split}$$

$$\begin{split} E(Q_{21}) &= S_{12} \times P(X_i = 2) \times P(Y_i = 1) \\ &= S_{12} \times G(\alpha_1, \beta_1, T_1) G(\alpha_2, \beta_2, T_2) \\ &\times {t_1 \choose 2} {t_2 \choose 1} \left(\frac{\Gamma(\alpha_1 + 2)\Gamma(t_1 + \beta_1 - 2)}{\Gamma(t_1 + \alpha_1 + \beta_1)} \right) \left(\frac{\Gamma(\alpha_2 + 1)\Gamma(t_2 + \beta_2 - 1)}{\Gamma(t_2 + \alpha_2 + \beta_2)} \right) \end{split}$$

並成立以下近似式:

$$\frac{E(Q_{12})}{E(Q_{21})} = \left(\frac{t_2 - 1}{t_1 - 1}\right) \left(\frac{t_1 + \beta_1 - 2}{t_2 + \beta_2 - 2}\right) \approx \left(\frac{t_2}{t_1}\right) \left(\frac{t_1 + \beta_1}{t_2 + \beta_2}\right)$$

由上推得:

$$\begin{cases} \beta_1 = \frac{E(Q_{12})}{E(Q_{21})} \left(\frac{t_1}{t_2}\right) (t_2 + \beta_2) - t_1 \\ \beta_2 = \frac{E(Q_{21})}{E(Q_{12})} \left(\frac{t_2}{t_1}\right) (t_1 + \beta_1) - t_2 \end{cases}$$

又可從式 (18) = 式 (22) 得:

$$\begin{split} &(t_1+\beta_1)^2 = \frac{E(Q_{11})}{E(Q_{22})} \frac{E(Q_{12})}{E(Q_{21})} t_1^2 \Rightarrow \beta_2^2 + 2t_2\beta_2 + t_2^2 \left(1 - \frac{E(Q_{11})}{E(Q_{22})} \frac{E(Q_{12})}{E(Q_{21})}\right) = 0 \\ &(t_2+\beta_2)^2 = \frac{E(Q_{11})}{E(Q_{22})} \frac{E(Q_{21})}{E(Q_{12})} t_2^2 \Rightarrow \beta_2^2 + 2t_2\beta_2 + t_2^2 \left(1 - \frac{E(Q_{11})}{E(Q_{22})} \frac{E(Q_{21})}{E(Q_{12})}\right) = 0 \end{split}$$

並依公式解 $\frac{-b\pm\sqrt{b^2-4ac}}{2a}$,得:

$$\begin{cases} \beta_1 = -t_1 + t_1 \sqrt{\frac{E(Q_{11})}{E(Q_{22})} \frac{E(Q_{12})}{E(Q_{21})}} = \left(-1 + \sqrt{\frac{E(Q_{11})}{E(Q_{22})} \frac{E(Q_{12})}{E(Q_{21})}} \right) t_1 \\ \beta_2 = -t_2 + t_2 \sqrt{\frac{E(Q_{11})}{E(Q_{22})} \frac{E(Q_{21})}{E(Q_{12})}} = \left(-1 + \sqrt{\frac{E(Q_{11})}{E(Q_{22})} \frac{E(Q_{21})}{E(Q_{12})}} \right) t_2 \end{cases}$$

最終得:

$$S_{wBB2} = D_{12} + E(Q_{00}) + E(Q_{0+}) + E(Q_{+0})$$

其中 $E(Q_{00})$ 等於 (14)、 $E(Q_{0+})$ 等於 (15) 且 $E(Q_{0+})$ 等於 (16),且在 (14) 中的 β_1 與 β_2 使用帶入 (25) 計算。

此外,在 Chao and Lin (2012) 中提出雨群集的取後不放回的共同種估計方式:

標準差估計

根據 $(Q_0,Q_1,...,Q_t)$ 的漸近分布,其服從大小為S以及機率為 $\left(\frac{E[Q_0]}{S},\frac{E[Q_1]}{S},...,\frac{E[Q_t]}{S}\right)$ 的多項分布 (multinomial distribution)。所提出的物種豐富度估計量的變異數估計量可以使用 delta 方法導出,表示為

$$\widehat{var}(\widehat{S_{12}}) = \sum_{i=1}^{t} \sum_{j=1}^{t} \frac{\partial \widehat{S_{12}}}{\partial Q_i} \frac{\partial \widehat{S_{12}}}{\partial Q_j} \widehat{cov}(Q_i, Q_j)$$

$$\not \perp + \widehat{cov}(Q_i, Q_j) = \begin{cases} Q_i \left(1 - \frac{Q_i}{\hat{S}}\right), & \text{if } i = j \\ -\frac{Q_i Q_j}{\hat{S}}, & \text{if } i \neq j \end{cases}$$
(37)

95%信賴區間

在此,物種豐富的信賴區間通過假 $\hat{S}_{12}-D_{12}$ 符合對數常態分佈 (log normal distribution) (Chiu et al., 2014),為此確保了信賴區間之下限值大於觀察到的物種豐富度。故,物種豐富度之 95%信賴區間為:

$$[D_{12} + \frac{\widehat{S_{12}} - D_{12}}{R}, D_{12} + (\widehat{S}_{12} - D_{12}) \times R]$$
 其中 $R = \left\{1.96 \left[\log\left(1 + \frac{\widehat{var}(\widehat{S_{12}})}{(\widehat{S_{12}} - D_{12})^2}\right)\right]^{\frac{1}{2}}\right\}$ 以此計算 95%信賴區間的樣本涵蓋率 (95% confidence interval coverage rate, 95% CI Coverage) 。