

兩種 Shannon 生物歧異度指數矯正估式之比較

楊滿霞¹、歐尚靈²、汪家琪³、彭雲明^{1*}

¹ 國立臺灣大學農藝系

² 國立中興大學農藝系

³ 致理技術學院財經系

摘要

Shannon 生物歧異度指數估值的偏誤 (bias) 矯正，一直以來都是生態統計學者關心的問題。這個問題在 2003 年 Chao and Shen 提出的估式後得到相當程度的解決，Pla 在 2004 年也提出了一個估式，我們以蒙地卡羅的模擬研究來比較這兩種估式的表現。研究結果顯示：(1) Chao and Shen 的估式在 CV 值 (相對豐量之變異係數) 低而且樣本數少的場合表現優於 Pla 的估式；(2) 此種優勢隨 CV 值增大逐漸減弱，樣本數增加至 100 以上時兩者表現相當接近；(3) 在大樣本時 Pla 估式的表現開始優於 Chao and Shen 的估式，原因是在大樣本時，偏誤幾乎消失，矯正的過程使得負向偏量變成正向偏量；(4) Pla 的估式其變異程度小於 Chao and Shen 的估式，此種優勢隨著樣本數變大而遞減；及(5) 為了建構有效的信賴區間，我們對於 Pla 所提的變方估值做了矯正，根據矯正後的變方估值所建構的信賴區間均能滿足 95% 的信賴係數。

關鍵詞：生物歧異度指數、偏誤矯正、靴環法、變異係數、信賴區間、涵蓋率。

Comparison of Two Shannon Index Estimators

* 通信作者, pong@ntu.edu.tw

投稿日期：2008 年 9 月 12 日

接受日期：2008 年 10 月 28 日

作物、環境與生物資訊 5:258-267 (2008)

Crop, Environment & Bioinformatics 5:258-267 (2008)

189 Chung-Cheng Rd., Wufeng, Taichung Hsien 41301, Taiwan ROC

Man-Hsia Yang¹, Shang-Ling Ou², Chia-Chi Wang³ and Yun-Ming Pong^{1*}

¹ Department of Agronomy, National Taiwan University, Taipei 10617, Taiwan ROC

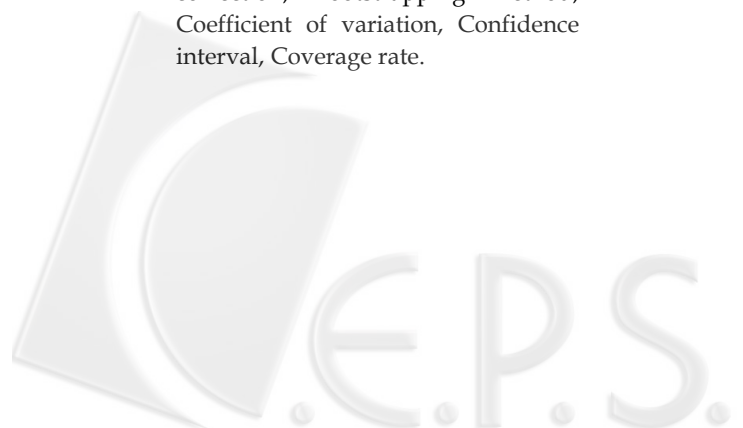
² Department of Agronomy, National Chung-Hsing University, Taichung 40227, Taiwan ROC

³ Department of Finance, Chihlee Institute of Technology, Banciao 22050, Taiwan ROC

ABSTRACT

The bias correction of Shannon index estimate had been a major concern for a long time. This problem was addressed by Chao and Shen (2003) and Pla (2004). The former is a theoretical approach and the latter is an empirical one. We did Monte Carlo simulations on these two approaches and examined if the simple intuitive approach by Pla was as good as the smart and sophisticated approach by Chao and Shen. The results of our study show that (1) Chao and Shen's estimator is superior when the coefficient of variation (CV) of relative abundance is low (less than 1), and sample sizes are small ($n \leq 75$); (2) the superiority is diminished for intermediate and high CV values (greater than 1) and under increasing sample sizes ($n \geq 75$); (3) the Pla estimator performs better than Chao and Shen's for large sample sizes ($200 \leq n \leq 400$) and high CV values; (4) Pla's estimator is less variable than that of Chao and Shen's; and (5) an adjusted variance estimator of Pla's biodiversity index estimator is proposed so that the coverage rate of confidence interval can achieve 95% of the nominal level.

Key words: Shannon's biodiversity index, Bias correction, Bootstrapping method, Coefficient of variation, Confidence interval, Coverage rate.



前言

在生態學的研究上，最常用來度量一群落 (community) 內物種歧異度的指標為 Shannon index，以符號 H 表示，此指標由 Shannon 在 1948 年提出，其計算式為

$$H = -\sum_{i=1}^S P_i \log P_i$$

此式中之 S 為群落內物種之總數， P_i 為第 i 個物種所佔的比率。由於 S 與 P_i 都是未知之參數，必須經由樣本去估算，此指標之最大概度估值為

$$\hat{H}_m = -\sum_{i=1}^{\hat{S}} \frac{X_i}{n} \log \left(\frac{X_i}{n} \right) \quad (1)$$

上式中之 \hat{S} 為樣本中所含之物種數， X_i 為樣本中第 i 個物種的個體數， n 為樣本中的總個體數。也就是說，以樣本中之總物種數估算群落中之總物種數，而且以樣本中第 i 個物種的出現比率估算群落中第 i 個物種所佔的比率。這個估算式的構想很直接，而且容易理解與計算，但是具有負向的偏誤，其偏誤已由 Basharin (1959) 導出

$$E(\hat{H}_m) = H - \frac{S-1}{2n} + O(n^{-2})$$

此式顯示樣本數 (n) 越大，偏誤則越小，但是對一特定樣本來說其偏誤由總物種數 S 與樣本數 n 決定。除了偏誤外，此估式另一個困擾是難以有效的估算其標準差。Manly (1997) 指出以靴環法 (bootstrapping) 的方式所得的信賴區間，對於真值 H 的涵蓋率低於名義上宣稱的 95%。這兩個困擾生態學者的問題由 Chao and Shen (2003) 與 Pla (2004) 提出修正的方法，前者的估式以 \hat{H}_c 表示，後者則以 \hat{H}_p 表示。前者 \hat{H}_c 經由理論的方式推導而得，後者 \hat{H}_p 則根據各種不同數值的歧異

度下群落結構模擬之數據求得。兩者均以電腦模擬研究的方式驗證其所提出之方法的有效性。本研究之目的在於探討 \hat{H}_c 與 \hat{H}_p 在三方面的表現之相似與相異之處：(1) 偏誤的矯正程度，(2) 變方的大小，(3) 標準差的估值。在比較兩個估式之前，先於下一節簡介此兩個估式。

材料與方法

在本節中我們將簡短介紹 Chao and Shen 與 Pla 兩者的估式。Chao and Shen (2003) 透過 Horvitz-Thompson 的方式將最大概度估式 \hat{H}_m 朝無偏誤的方向矯正為

$$\hat{H}_c = -\sum_{i=1}^{\hat{S}} \frac{P_i \log(P_i)}{1 - (1 - P_i)^n} I(A_i)$$

式中之 $I(A_i)$ 是一個 A_i 的指示函數， A_i 是一個事件代表第 i 個物種是否出現在樣本中，若是出現則 $I(A_i) = 1$ ，反之 $I(A_i) = 0$ ； $1 - (1 - P_i)^n$ 是第 i 個物種被抽取到的機率。其次將最大概度估式 \hat{H}_m 中之相對豐量 (relative abundance) $\hat{P}_i = X_i / n$ 修正為 $\hat{C}\hat{P}_i$ ，式中之 \hat{C} 為樣本遮蓋率，修正的用意在於將沒取到的物種考慮進來。修正後的 \hat{H}_c 具有下列型式

$$\hat{H}_c = -\sum_{i=1}^{\hat{S}} \frac{\hat{C}\hat{P}_i \log(\hat{C}\hat{P}_i)}{1 - (1 - \hat{C}\hat{P}_i)^n} I(A_i)$$

上式中之 $\hat{C} = 1 - f_1 / n$ ， f_1 為樣本中個體數為 1 的物種數， n 為樣本中之總個體數。修正後的 \hat{H}_c 與修正前的 \hat{H}_m 其外型基本上是一致的。若是將樣本資料整理成下列簡潔的方式， f_m 為樣本中具有 m 個體數的物種數。

m	1	2	...	k	...	n
f_m	f_1	f_2	...	f_k	...	f_n



則上述之 \hat{H}_c 可以改寫成配合簡潔樣本資料的型式

$$\hat{H}_c = -\sum_{k=1}^n f_k \frac{\left[\left(1 - \frac{f_1}{n}\right) \frac{k}{n} \right] \log \left[\left(1 - \frac{f_1}{n}\right) \frac{k}{n} \right]}{1 - \left[1 - \left(1 - \frac{f_1}{n}\right) \frac{k}{n} \right]^n}$$

Chao and Shen (2003) 也導出對應於此式之大樣本變方為

$$\text{var}(\hat{H}_c) = \sum_{k=1}^n \sum_{j=1}^n \frac{\partial \hat{H}_c}{\partial f_i} \frac{\partial \hat{H}_c}{\partial f_j} \text{cov}(f_i, f_j)$$

式中之 $\text{cov}(f_i, f_j) = -f_i f_j / \tilde{S}$ ，當 $i \neq j$ ， $\text{cov}(f_i, f_i) = f_i(1 - f_i) / \tilde{S}$ ，此處之 \tilde{S} 為 Chao and Lee (1992) 以及 Chao *et al.* (2000) 導出的物種數估式。

Pla (2004) 的出發點則是傳統的靴環法 (Efron 1979)，亦即透過靴環的方式從樣本再取樣，並按照式(1)計算，求得一序列之靴環估值 $H_1^*, H_2^*, H_3^*, \dots, H_n^*$ 。其次考慮歧異度真值 H 與最大概度估值 \hat{H}_m 的偏誤估值 $\hat{H}_m - H$ ，此估值可分割成兩部分

$$\hat{H}_m - H = (\hat{H}_m - \bar{H}^*) + (\bar{H}^* - H)$$

式中之 \bar{H}^* 為靴環估值均數。上式等號右邊第一項 $(\hat{H}_m - \bar{H}^*)$ 是可從樣本取得之數值，第二項之數值 $(\bar{H}^* - H)$ 則無法從樣本取得，因為 H 是未知的群落歧異度。由於偏誤隨著總物種數增加而負向增大，因為偏誤的理論值為 $\text{bias} = -(S-1)/n$ ，所以 Pla 構想一個複回歸模式

$$(\hat{H}_m - H) = \beta_1(\hat{H} - \bar{H}^*) + \beta_2\bar{H}^* + \beta_3\hat{S} + \delta$$

來描述偏誤的幅度。Pla 模擬各種不同形式的群落物種結構，涵蓋廣闊的歧異度值 ($H = 0.557 \sim 4.097$)，並在此情況下進行抽樣以計算 \hat{H}_m 與 \hat{S} ，以及再抽樣以計算靴環估值

均數 \bar{H}^* 。然後做複回歸模式之配適，得到下列估算偏誤的複回歸式

$$(\hat{H}_m - H) = 1.73(\bar{H}^* - \hat{H}_m) + 0.02\bar{H}^* - 0.0003\hat{S} + \delta$$

將此式改寫後即得 Pla 的估式

$$\hat{H}_p = 2.73\hat{H}_m - 1.75\bar{H}^* + 0.0003\hat{S}$$

至於 \hat{H}_p 的變方 Pla 也是推薦靴環法的方式估計 $\sigma_{\hat{H}_p}^2$ ，但是這個變方估值有低估的傾向，無法有效的估計 $\sigma_{\hat{H}_p}^2$ ，因此 Yang (2006) 提出下列矯正式

$$\sigma_{\hat{H}_p}^2 = \frac{\sigma_{\hat{H}^*}^2}{0.46}$$

上式中之 $\sigma_{\hat{H}^*}^2$ 為靴環法所求出的物種歧異度估值的變方。

我們將以模擬的方式比較 \hat{H}_c 與 \hat{H}_p 兩者對於偏誤的矯正程度以及其變方的大小。在進行模擬比較前，我們沿用 Chao and Shen 的方式定義物種相對豐量的變異係數如下

$$\text{CV} = \frac{\sum_{i=1}^S [(P_i - \bar{P})^2 / S]^{1/2}}{\bar{P}}$$

式中之 $\bar{P} = \sum_{i=1}^S P_i / S$ 。此數值是用來描述各種型式的群落結構，我們將在各種 CV 數值下觀察 \hat{H}_c 與 \hat{H}_p 的表現。參考 Chao and Shen 的模擬架構，以四種物種豐富度模式進行模擬

1. $P_i = cx_i$ ，式中之 (x_1, x_2, \dots, x_S) 為從均勻分布 $U(0,1)$ 中取出之一組隨機樣本。
2. $P_i = cx_i$ ，式中之 (x_1, x_2, \dots, x_S) 為從平均值為 1 的指數分布中取出之一組隨機樣本。
3. $P_i = c / (i+2)$ ， $i = 1, 2, \dots, S$ 。
4. $P_i = c / i$ ， $i = 1, 2, \dots, S$ 。

模擬研究分成兩個層面，第一個層面是在比較廣泛的 CV 值下觀察兩者的表現。為了達到這個目的，我們從每個模式中取出 91



個群落，群落所含的物種數 S 從 10 至 100，以期所選出之群落代表各種不同的 CV 值與 H 值。從每個群落取出 100 個樣本，按照此方式每個模式可評估 91 個偏誤值，四種模式下則有 364 個偏誤值。同樣的方式可求得 364 個變方值與變方估值。第二個層面是縮小範圍，僅從四種特定的 CV 值下觀察兩者的表現。

結果與討論

我們將分成三個小節來說明比較的結果，(1)偏誤的矯正，(2)變方的大小，及(3)變方估值的正確性。

1. 各種 CV 值下偏誤的矯正

將 \hat{H}_c 與 \hat{H}_p 的偏誤值各 364 個與其對應之 CV 值繪成散布圖如 Fig. 1 所示。這是在樣本數 $n = S$ 下所得的結果，也就是樣本數為一倍的物種數。Fig. 1 的橫軸為 CV 值，此數值小表示群落中所含物種的相對豐量較均勻，在此情形下，物種歧異度 H 較大，此時群落中之物種數出現於樣本的比率較高，因此偏誤較小， \hat{H}_c 與 \hat{H}_p 的偏誤矯正效果無法顯現；但是在 CV 值大時，表示相對豐量非常不均勻，豐量大的物種多數出現於樣本，但是稀有物種出現在樣本的機率低。因此，有相當數目的物種不會出現在隨機樣本，進而造成較大的偏誤，此種場景正是評估 \hat{H}_c 、 \hat{H}_p 兩者對於偏誤矯正的能力。Fig. 1 中之符號 \square ， \times 顯示：雖然經過矯正，負向的偏誤仍然存在，但是幅度已經小至 0.15 以內， \hat{H}_c 的矯正稍微優於 \hat{H}_p ，前者在 -0.05 左右，後者在 -0.10 左右，兩者相差之幅度約為 0.05。若是樣本數增加為 $n = 3S$ 時其偏誤值矯正情形如 Fig. 2 所示。Fig. 2 中之符號 \times 顯示出 \hat{H}_c 的偏誤隨著 CV 值增大而由負值變成正值， \square 則顯示 \hat{H}_p 的趨勢則相反，由正值下降至無偏的狀態。對於更大的樣本數所得的結果，例如 $n = 5S$ 本文中雖無顯示，但此種現象依然維持著，因此我們推測：大樣本似乎是對 \hat{H}_p 的

偏誤矯正較為有利。

2. 各種 CV 值下的誤差均方的大小

這一小節是以模擬的方式估計 $\sigma_{H_c}^2$ 以及 $\sigma_{H_p}^2$ 我們在 364 個模擬群落下每一群落取出 100 樣本，根據樣本估值 $\hat{H}_i, i = c, p$ 以下列方式估算其所對應的誤差均方，為了方便起見我們仍用變方的符號 σ^2 來表示，而且以「變方」來稱呼「誤差均方」。

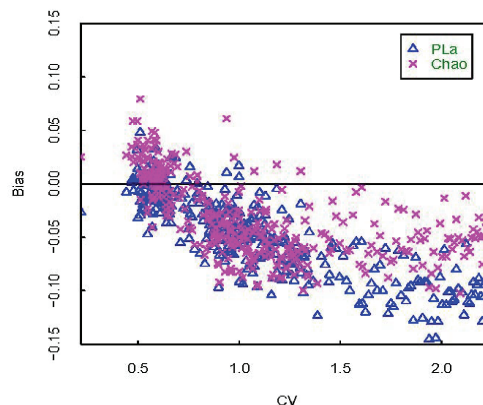


Fig. 1. Scatterplot of bias of \hat{H}_c and \hat{H}_p under various CV's when sample size $n = S$. Here \hat{H}_c and \hat{H}_p are biodiversity index estimators of Chao and Pla, respectively.

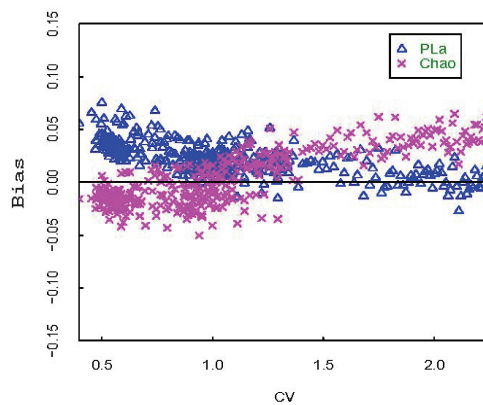


Fig. 2. Scatterplot of bias of \hat{H}_c and \hat{H}_p under various CV's when sample size $n = 3S$.



$$\hat{\sigma}_{\hat{H}_i}^2 = \frac{1}{100-1} \sum_{h=1}^{100} (\hat{H}_{ih} - \bar{H}_i)^2 \quad i=c,p$$

Fig. 3 是樣本數為一倍時 ($n=S$)， \hat{H}_c 、 \hat{H}_p 兩種歧異度估式的變方隨著 CV 值而變動的圖示。符號 \square ， \times 出現的位置顯示兩者大致重疊，因此對於 \hat{H}_c 、 \hat{H}_p 兩者來說不易察覺何者具有優勢，變方的數值在 0.14 以下。在樣本數增加為三倍的物種數時 $n=3S$ ，兩個估式的變方均變小在 0.04 以下，如 Fig. 4 所示。我們將在下列兩小節中，在特定的 CV 值下觀察何者較具有優勢。

3. 固定 CV 值下的偏誤與誤差均方根

在這小節，我們將焦點限制在四個不同的 CV 值下，觀察 Chao 與 Pla 兩者的估值偏誤與誤差均方根 (root mean square error; RMSE) 隨著樣本數的變化。四種不同的 CV 值: 0.547、0.920、1.343、2.253 分別代表前述四種模式相對豐量的均勻程度，數值小者均勻度高，數值相對較大者表示相對豐量較不均勻。在四個模式與三種樣本數的 12 種組合下，以電腦模擬的方式抽取 1,000 個隨機樣本，計算樣本估值 \hat{H}_c 與 \hat{H}_p ，並觀察其偏誤與誤差均方根之數值。模擬的結果顯示於 Fig. 5。此圖顯示: (1) \hat{H}_c 在 CV 值小 (Model 1) 的時後表現最佳，其偏誤幾乎不存在，隨著 CV 漸增 (Models 2,3,4)，偏誤為負值而且幅度加大，在樣本數小的時候 ($n=50$) 偏誤相對較大，但是隨樣本數增加至 $n=100$ 時，偏誤即逐漸縮小。(2) \hat{H}_p 在 $n=50$ 時雖不如 \hat{H}_c ，但是隨著樣本數增加，其偏誤亦逐漸減少而接近 \hat{H}_c 的表現。其次將 \hat{H}_c 與 \hat{H}_p 兩者的誤差均方根繪成圖形，如 Fig. 6。圖形顯示 \hat{H}_c 的誤差均方根在前三種模式的表現均較 \hat{H}_p 稍大，隨著 CV 值加大，兩者的誤差均方根逐漸接近。

4. 大樣本下的偏誤 $n=100, 200, 300, 400$

在樣本數小於 100 時， \hat{H}_c 的偏誤矯正的表现均優於 \hat{H}_p ，但是當樣本數增加時，情

況卻逐漸逆轉。Fig. 7 顯示當樣本數逐漸增大時 \hat{H}_c 與 \hat{H}_p 兩者之偏誤改變之情形。Fig. 7 中的四個子圖與 Fig. 2 的結果 ($n=3S$) 相當一致。當 CV 值較大時 (Model 4)， \hat{H}_c 似乎是過度矯正了，因為此時單純的最大概度估值 \hat{H}_m 的偏誤都已經大幅減少，其幅度從 0.5 ($n=100$) 降至 0.15 ($n=400$)。對於 \hat{H}_p 來說， $n=200$ 時似乎是關鍵點，對於更大的樣本數 Pla 的估式具有優勢。

5. 信賴區間的覆蓋率

除了上述最重要的主題：偏誤與變異大小之外，另一個主題就是有關於變異之估算。以往藉由靴環法估算最大概度估值 \hat{H}_m

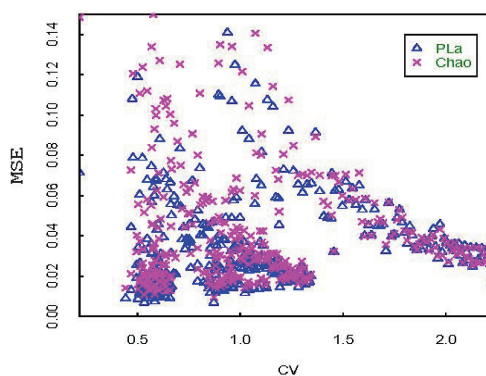


Fig. 3. Scatterplot of MSE of \hat{H}_c and \hat{H}_p under various CV's when sample size $n = S$.

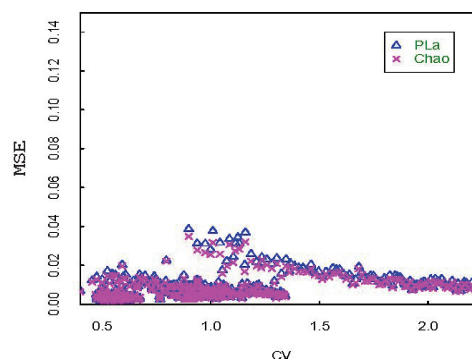


Fig. 4. Scatterplot of MSE of \hat{H}_c and \hat{H}_p under various CV's when sample size $n = 3S$.

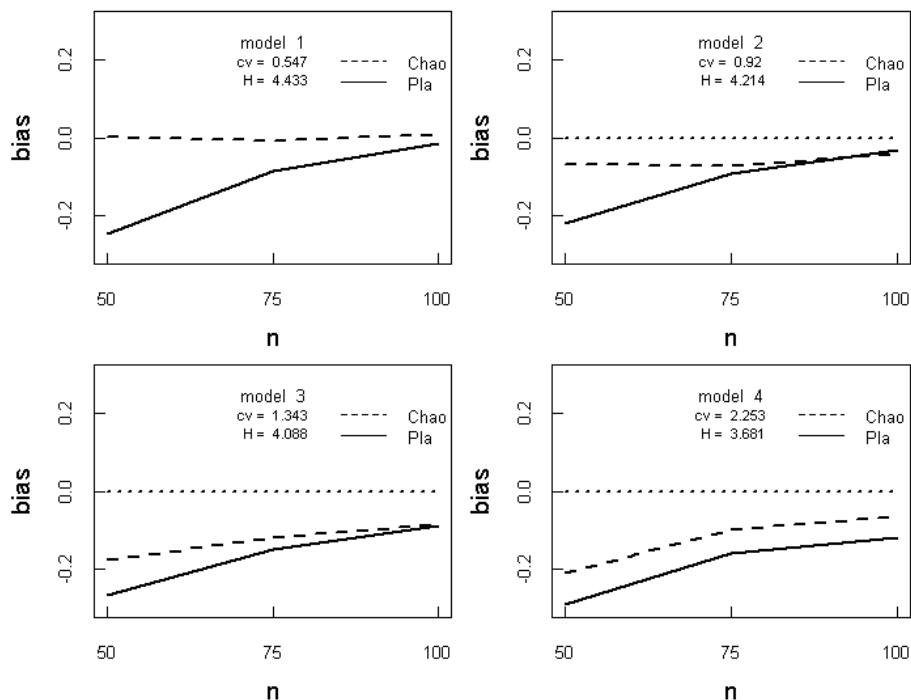


Fig. 5. Bias curves of \hat{H}_c and \hat{H}_p under four models when sample size $n \leq 100$. Here \hat{H}_c and \hat{H}_p are biodiversity index estimators of Chao and Pla, respectively.

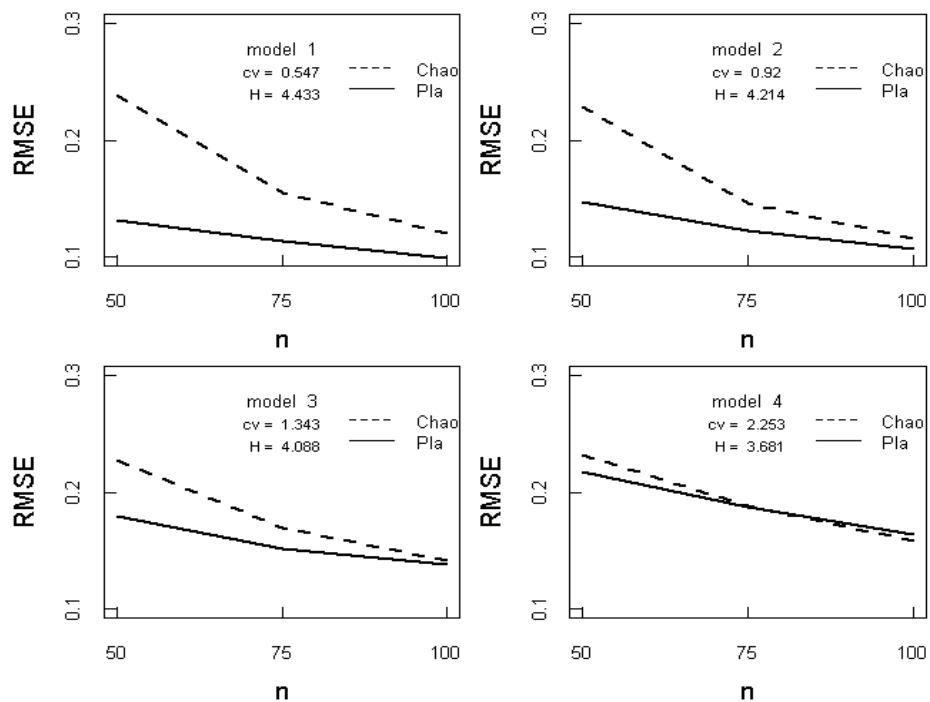
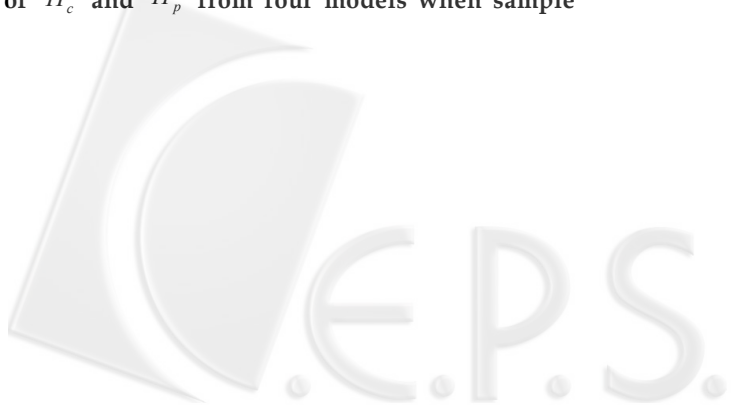


Fig. 6. RMSE (root mean squared error) curves of \hat{H}_c and \hat{H}_p from four models when sample size $n \leq 100$.



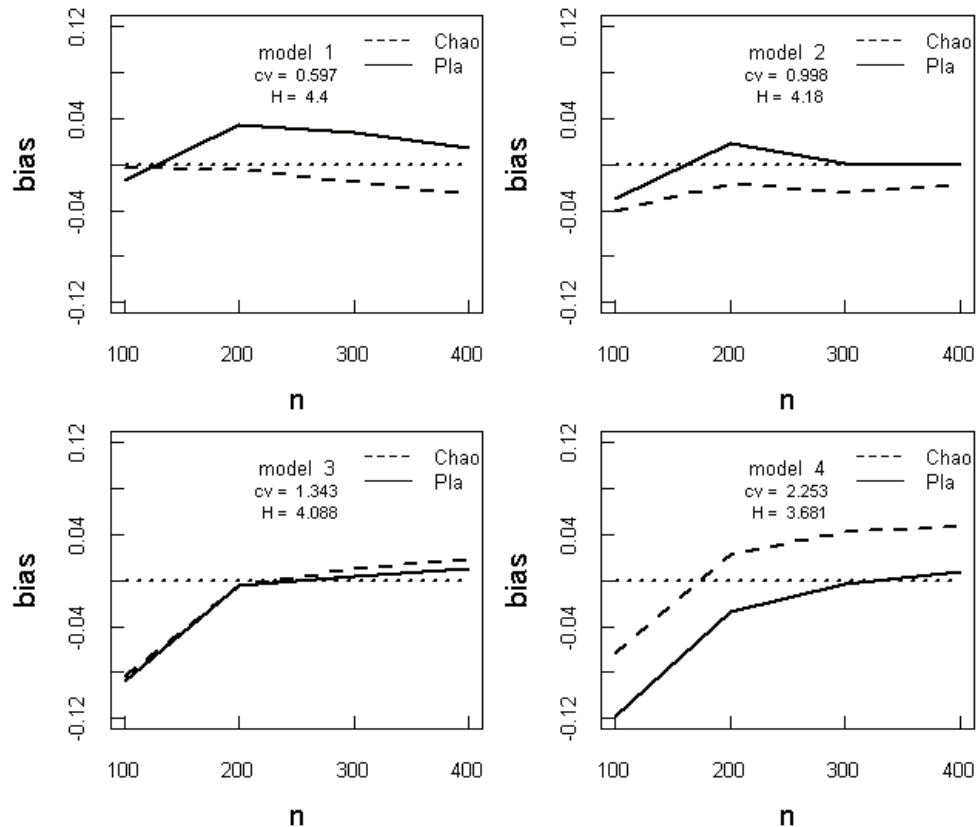


Fig. 7. Bias curves of \hat{H}_c and \hat{H}_p under four models when sample size $n \leq 100$.

之標準差，其最大的問題為低估。這個問題同樣發生在 Pla 所提出之估式 \hat{H}_p 低估標準差的後果就是對於 H 的信賴區間無法有效的估算，因為覆蓋率偏低無法達到名義上的 95%。Yang (2006) 在其碩士論文提出一矯正式

$$\hat{\sigma}_{\hat{H}_p}^2 = \frac{1}{0.46} \cdot \hat{\sigma}_{\hat{H}}^2$$

式中之 $\hat{\sigma}_{\hat{H}_p}^2$ 為 \hat{H}_p 之變方估值，亦即前述之誤差均方(MSE)由 1,000 個隨機樣本所得的 \hat{H}_p 計算而得，由於偏誤已經大幅改善，因此將誤差均方視為變方之估值； $\hat{\sigma}_{\hat{H}}^2$ 則為每一隨機樣本所得的變方估值，是透過靴環法得到的。Fig. 8 顯示這個矯正式大致上能有效的描述此兩種不同方式的變方估值。根據此矯正式所得的標準差建立的信賴區間大致上已達

到名義上宣稱的 95% 信賴係數。Table 1 顯示：當樣本數為 $n = 3S$ 時，覆蓋率均已超過 95%，這個顯示變方矯正式有過度矯正的趨勢，這個趨勢顯示在 Fig. 8 的左下角，當樣本數變大時， \hat{H}_p 之變方變小，此時縱軸之數值多數大於橫軸數值，因此對應點位於參考線(對角線)之上方。

6. 數例與結論

綜合上述之結果，我們得到下列之結論：(1) 使用 Chao 的矯正式最有利的情境是 CV 值低而且樣本數小的時候，例如在模式一與模式二，隨著 CV 值增大 Chao 的估式偏誤增加，雖然偏誤會隨著樣本數增加而減少，但是樣本數增大至 $n = 200$ 時偏誤變為正值，而且亦緩緩上升。(2) Pla 的估式表現則在於 CV 值較大而且樣本數較大時較 Chao 為優。



我們以七組實際的數據(列於 Table 2)說明歧異度指數之估計，其結果列於 Table 3，並透過先前模擬研究的結果來解讀這些估值。對於 Lee, Janzan I, Janzan II, Batten I Batten II 五組數據，其中後四者在 Chao and Shen

(2003) 已有討論、這五者是 \hat{H}_c 最具優勢的情境，我們在 Table 3 可以看出 \hat{H}_p 的數值與 \hat{H}_c 非常接近。對於 Hwang I, Hwang II 兩組數據，由於樣本數較大， \hat{H}_m 已經接近 \hat{H}_p ，但是 \hat{H}_c 的優勢已不再，反而有高估的傾向。

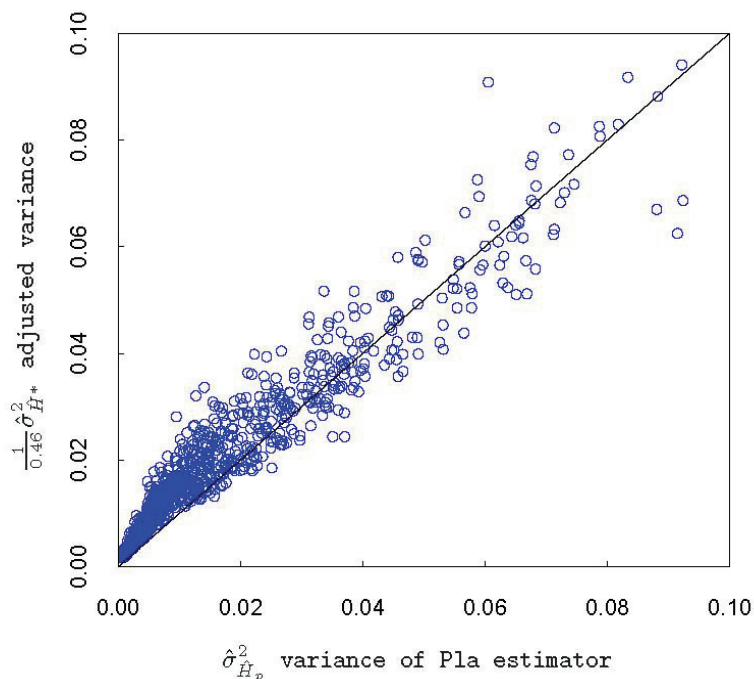


Fig. 8. Scatter plot of adjusted variance estimate $\frac{1}{0.46} \hat{\sigma}_{\hat{H}^*}^2$ versus $\hat{\sigma}_{\hat{H}_p}^2$ which is estimated by 1,000 sample estimates of \hat{H}_p . The diagonal line is the reference line $\frac{1}{0.46} \hat{\sigma}_{\hat{H}^*}^2 = \hat{\sigma}_{\hat{H}_p}^2$.

Table 1. Coverage rates under 36 scenarios. Each scenario is a combination of a model with a specified CV value, a sample size, and a specified number of species.

Sample Size n	Number of species S	model			
		1	2	3	4
n=S	10	91.8	94.9	93.8	94.5
	50	97.1	94.8	94.6	91.3
	100	96.4	94.6	91.0	92.2
n=2S	10	97.5	97.3	97.4	95.3
	50	98.0	96.8	96.7	97.2
	100	95.2	97.3	97.0	97.8
n=3S	10	98.2	96.3	99.1	97.9
	50	98.2	98.5	98.1	98.2
	100	97.0	98.1	98.1	97.8

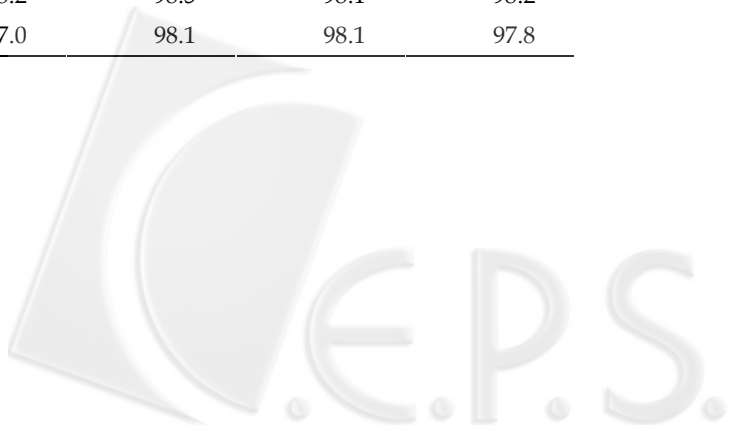


Table 2. Seven data sets for illustrative purpose.

Lee's (2005) snake data											
m	1	2	3	4	5	13	16				
fm	3	3	2	5	1	1	1				
Janzan's (1973, cf Chao 2003) insect data (day)											
m	1	2	3	4	5	6	11				
fm	59	9	3	2	2	2	1				
Janzan's (1973, cf Chao 2003) insect data (night)											
m	1	2	3	5	7	10	14	16	18		
fm	56	9	7	2	1	1	1	1	1	1	
Batten's (1976, cf Magurran, 1988) bird bata I											
m	1	2	3	5	6	11	16	21	25	26	35
fm	4	3	5	1	1	1	1	1	1	1	1
Batten's (1976, cf Magurran, 1988) bird bata II											
m	1	2	3	4	5	9	11	14	20	30	65
fm	2	1	2	1	1	1	1	1	1	2	1
Hwang's (2000) butterfly data I											
m	1	2	3	4	6	7	10	17	19	24	27
fm	7	5	1	2	3	2	2	1	1	1	1
m	28	36	39	44	73	76	113				
fm	1	1	1	1	1	1	1				
Hwang's (2000) butterfly data II											
m	1	2	3	4	5	8	9	10	13	14	17
fm	9	3	2	1	1	1	1	1	1	1	1
m	22	25	83	92							
fm	1	1	2	1							

Table 3. Estimate and associated SE (in parenthesis) of Shannon index obtained from \hat{H}_m , \hat{H}_c and \hat{H}_p .

Source	CV	\hat{S}	n	MLE	Chao	Pla
Lee (snake)	0.944	16	69	2.434 (0.099)	2.562 (0.115)	2.608 (0.165)
Janzan I (insect)	0.959	78	127	4.077 (0.074)	4.699 (0.211)	4.621 (0.126)
Janzan II (insect)	0.953	79	170	3.832 (0.088)	4.297 (0.205)	4.230 (0.139)
Batten I (bird)	1.183	20	170	2.408 (0.071)	2.491 (0.101)	2.466 (0.109)
Batten II (bird)	1.414	14	198	2.056 (0.066)	2.093 (0.159)	2.080 (0.097)
Hwang I (butterfly)	1.472	33	576	2.703 (0.043)	2.746 (0.136)	2.708 (0.065)
Hwang II (butterfly)	1.627	27	406	2.311 (0.057)	2.378 (0.085)	2.333 (0.158)



引用文獻

- Bashrain GP (1959) On a statistical estimate for the entropy of a sequence of independent random variables. **Theory Prob. Its Appl.** 4:333-6.
- Chao A, WH Hwang, YC Chen, CY Kuo (2000) Estimating the number of shared species in two communities. **Statist. Sinica** 10:277-246.
- Chao A, SM Lee (1992) Estimating the number of classes via sample coverage. **JASA** 87: 210-217.
- Chao A, TJ Shen (2003) Nonparametric estimation of Shannon's index of diversity when there are unseen species in sample. **Environ. Ecol. Statist.** 10:429-443.
- Efron B (1979) Bootstrap Methods: Another Look at the Jackknife. **Statist.** 7 (1): 1-26.
- Hwang S (2000) A study of butterfly diversity in two area of Kinmen Island. http://www.kmnp.gov.tw/chinese/file/kmnp_research89_insect.pdf
- Lee TH (2005) Ecological patterns of distribution on gradients of elevation and species diversity of snakes in southern Taiwan. **Amphibia Reptilia** 26:325-332.
- Magurran AE (1988) *Ecological Diversity and Its Measurement*. p.145-146. Princeton University Press, New York.
- Manly BFJ (1997) *Randomization, Bootstrap and Monte Carlo Methods in Biology*. 2nd Ed. Chapman & Hall, London.
- Pla L (2004) Bootstrap confidence intervals for the Shannon biodiversity index: A simulation study. **JABES** 9: 42-56.
- Shannon, CE (1948) A mathematical theory of communication. **Bell Syst. Tech. J.** 27: 379-423, 623-656
- Yang MS (2006) A Simulation Study of Two Biodiversity Index Estimation Procedures-In Comparing Bias Correction and Confidence Interval Coverage Rate. Master Thesis. Department of Agronomy, National Taiwan University, Taipei.

