On the asymptotic variance of the Chao estimator for species richness estimation

Chang Xuan Mao, Sijia Zhang, Zhilin Liao

School of Statistics and Management,

Shanghai University of Finance and Economics, Shanghai, China

Supplementary Material

Proposition 1. *As s goes to infinity, $(\hat{\nu}-\nu)/\sqrt{s}$ converges weakly to a normal distribution with mean zero and variance $\gamma^2 = \gamma_1^2 + \gamma_2^2$, where*

$$\gamma_1^2 = \frac{p_1^2}{2p_2} + \frac{p_1^3}{p_2^2} + \frac{p_1^4}{4p_2^3}, \qquad \gamma_2^2 = \left\{1 - p_0 + \frac{p_1^2}{2p_2}\right\} \cdot \left\{p_0 - \frac{p_1^2}{2p_2}\right\}.$$

Proof of Proposition 1. With $\boldsymbol{m} = (n_0, n_1, n_2)^\top$ and $\boldsymbol{p} = (p_0, p_1, p_2)^\top$, as s goes to infinity, $\sqrt{s}(\boldsymbol{m}/s - \boldsymbol{p})$ converges weakly to $\mathcal{N}(\boldsymbol{0}, \mathrm{diag}(\boldsymbol{p}) - \boldsymbol{p}\boldsymbol{p}^\top)$, where $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$ is a multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix Σ and $\mathrm{diag}(\cdot)$ builds a diagonal matrix with its arguments as diagonal entries. Let $f(\boldsymbol{p}) = 1 - p_0 + p_1^2/(2p_2)$. By the delta method, as s goes to infinity, $(\hat{\nu}-\nu)/\sqrt{s} = \sqrt{s}(f(\boldsymbol{m}/s) - f(\boldsymbol{p}))$ converges weakly to $\mathcal{N}(0, \gamma^2)$, where $\gamma^2 = \nabla^\top f(\boldsymbol{p})\{\mathrm{diag}(\boldsymbol{p}) - \boldsymbol{p}\boldsymbol{p}^\top\}\nabla f(\boldsymbol{p})$ and $\nabla^\top f(\boldsymbol{p}) = (-1, p_1/p_2, -p_1^2/(2p_2^2))$. By some algebra, $\gamma^2 = \gamma_1^2 + \gamma_2^2$. □

Proposition 2. *The estimator $\tilde{\nu}$ is unbiased in the limit in the sense that*

$$E(\tilde{\nu}/s) = 1 - p_0 + p_1^2/(2p_2) + O\big((1-p_2)^s\big). \tag{1}$$

Proposition 3. *The $\tilde{\sigma}^2$ is unbiased in the limit in the sense that*

$$E(\tilde{\sigma}^2/s) = \gamma_1^2 + O\big(s^2(1-p_2)^s\big). \tag{2}$$

Proofs of Propositions 2 and 3. We will show that the following hold, i.e.,

$$E(\tilde{\nu}/s) = 1 - p_0 + p_1^2/(2p_2) - p_1^2/\{2p_2(1-p_2)\} \cdot (1-p_2)^s, \tag{3}$$

$$E(\tilde{\sigma}^2/s) = \gamma_1^2 - (1-p_2)^{-3} \cdot r(s, p_1, p_2)(1-p_2)^s. \tag{4}$$

For $i = 1$, 2 and 3, write

$$\begin{aligned} E\left\{\frac{n_1}{s}\prod_{j=1}^{i}\frac{(n_1-j)}{(n_2+j)}\right\} &= \sum_{n_1+n_2+u=s}\frac{n_1}{s}\prod_{j=1}^{i}\frac{(n_1-j)}{(n_2+j)} \cdot \frac{s!p_1^{n_1}p_2^{n_2}q^u}{n_1!n_2!u!} \\ &= \frac{p_1^{i+1}}{p_2^i}\sum_{n_1+n_2+u=s, n_1\geqslant i+1}\frac{(s-1)!p_1^{n_1-i-1}p_2^{n_2+i}q^u}{(n_1-i-1)!(n_2+i)!u!} \\ &= \frac{p_1^{i+1}}{p_2^i}\left\{1 - \sum_{j=0}^{i-1}\tbinom{s-1}{j}p_2^j(1-p_2)^{s-1-j}\right\}, \end{aligned}$$

where $u = s - n_1 - n_2$ and $q = 1 - p_1 - p_2$. Clearly, (3) and (4) hold. □

Proposition 4. *Both $\check{\nu}$ and $\check{\sigma}^2$ are unbiased in the limit, i.e.,*

$$E(\check{\nu}/s) = 1 - p_0 + p_1^2/(2p_2) + b_1(p_1, p_2)/s + O((1-p_2)^s), \tag{5}$$

$$E(\check{\sigma}^2/s) = \gamma_1^2 + b_2(p_1, p_2)/s + O(s^{-3/2}), \tag{6}$$

where $b_1(p_1, p_2) = p_1/(2p_2)$ and

$$b_2(p_1, p_2) = \sum_{i=1}^{3}\frac{\{p_2 + (p_1+p_2)(i-1)/2\}ip_1^i}{\{1 + I(i=1) + 3I(i=3)\}p_2^{i+1}}.$$

Proof of Proposition 4 . Write

$$E\left(\frac{n_1}{n_2+1}\right) = \sum_{n_1+n_2+u=s}\frac{n_1}{n_2+1} \cdot \frac{s!}{n_1!n_2!u!}p_1^{n_1}p_2^{n_2}q^u = \frac{p_1}{p_2}\Big\{1 - (1-p_2)^s\Big\}.$$

Note that (5) holds as, with $n_1^2 = n_1(n_1-1)+n_1$,

$$
\begin{aligned}
E\left(\frac{\check{\nu}-n}{s}\right) &= \frac{p_1^2}{2p_2}\left\{1-(1-p_2)^{s-1}\right\} + \frac{1}{s}\cdot\frac{p_1}{2p_2}\left\{1-(1-p_2)^{s}\right\} \\
&= \frac{p_1^2}{2p_2} + \frac{1}{s}\cdot\frac{p_1}{2p_2} - \left\{\frac{p_1^2}{2p_2(1-p_2)} + \frac{1}{s}\cdot\frac{p_1}{2p_2}\right\}(1-p_2)^s. \qquad \square
\end{aligned}
$$

To prove (6), we write

$$
\frac{\check{\sigma}^2}{s} = \frac{1}{2}\cdot\frac{(n_1/s)^2}{(n_2/s+1/s)} + \frac{(n_1/s)^3}{(n_2/s+1/s)^2} + \frac{1}{4}\cdot\frac{(n_1/s)^4}{(n_2/s+1/s)^3}. \tag{7}
$$

By the linearity of the expectation functional, one can consider the three terms in (7) one by one. Let $g_m(x_1,x_2) = x_1^{m+1}/x_2^m$, $m = 1$, 2 and 3, $x_1 \geqslant 0$ and $x_2 > 0$. Let f be one of g_1, g_2 and g_3, and

$$
f'_{i,1-i} = \frac{\partial f}{\partial x_1^i \partial x_2^{1-i}},\ f''_{i,2-i} = \frac{\partial^2 f}{\partial x_1^i \partial x_2^{2-i}},\ f'''_{i,3-i} = \frac{\partial^3 f}{\partial x_1^i \partial x_2^{3-i}}.
$$

Let $W_1 = n_1/s$ and $W_2 = (n_2+1)/s$. Note that $Ef(W_1,W_2) - f(p_1,p_2) = A_1 + A_2$, where, using Taylor expansion of $f(W_1,W_2)$ at (p_1,p_2) and the linearity of the expectation, and with ξ_i being between p_i and W_i, $i = 1$, 2,

$$
\begin{aligned}
A_1 &= f'_{1,0}E(W_1-p_1) + f'_{0,1}E(W_2-p_2) + 2^{-1}f''_{2,0}E(W_1-p_1)^2 \\
&\quad + 2^{-1}f''_{0,2}E(W_2-p_2)^2 + f''_{1,1}E(W_1-p_1)(W_2-p_2), \\
A_2 &= \sum_{i=0}^{3}\frac{1}{3!}\binom{3}{i}E\left\{f'''_{i,3-i}(\xi_1,\xi_2)(W_1-p_1)^i(W_2-p_2)^{3-i}\right\},
\end{aligned} \tag{8}
$$

where $f'_{1,0}$, $f'_{0,1}$, $f''_{2,0}$, $f''_{0,2}$ and $f''_{1,1}$ are evaluated at (p_1,p_2). One has

$$
\begin{aligned}
A_1 &= \frac{f'_{0,1} + 2^{-1}f''_{2,0}p_1(1-p_1) + 2^{-1}f''_{0,2}p_2(1-p_2) - f''_{1,1}p_1p_2}{s} + \frac{f''_{0,2}}{2s^2} \\
&= \frac{(mp_1^m/p_2^{m+1})\{p_2 + (p_1+p_2)(m-1)/2\}}{s} + \frac{m(m+1)p_1^{m+1}}{2s^2p_2^{m+2}}.
\end{aligned}
$$

Next we will seek an upper bound of $|A_2|$ by considering the terms in A_2 one by one. By the Cauchy-Schwarz inequality, one has

$$
\begin{aligned}
&\left|E\left\{f'''_{i,3-i}(\xi_1,\xi_2)(W_1-p_1)^i(W_2-p_2)^{3-i}\right\}\right|^2 \\
&\leqslant E\left|f'''_{i,3-i}(\xi_1,\xi_2)\right|^2 \cdot E\{(W_1-p_1)^i(W_2-p_2)^{3-i}\}^2.
\end{aligned} \tag{9}
$$

With $Z_{i,s} = (n_i - sp_i)/\{sp_i(1-p_i)\}^{1/2}$, by the Cauchy-Schwarz inequality,

$$E\{(W_1-p_1)^i(W_2-p_2)^{3-i}\}^2 \leqslant \Big\{E\{(W_1-p_1)^{4i}\}E\{(W_2-p_2)^{4(3-i)}\}\Big\}^{1/2},$$

$$E\{(W_1-p_1)^{4i}\} = s^{-2i}\{p_1(1-p_1)\}^{2i}EZ_{1,s}^{4i},$$

$$\begin{aligned} E\{(W_2-p_2)^{4(3-i)}\} &= s^{4(i-3)}\sum_{j=0}^{4(3-i)}\gamma_{ij}E(n_2-sp_2)^j \\ &= s^{2(i-3)}\left\{\sum_{j=0}^{4(3-i)-1}\gamma_{ij}\frac{\{p_2(1-p_2)\}^{j/2}EZ_{2,s}^j}{s^{2(3-i)-j/2}} + \{p_2(1-p_2)\}^{2(3-i)}EZ_{2,s}^{4(3-i)}\right\}, \end{aligned}$$

where $\gamma_{i,j} = \binom{4(3-i)}{j}$. Let μ_i denote the ith moment of $\mathcal{N}(0,1)$. Because the moment generating function of $Z_{i,s}$ converges to that of $\mathcal{N}(0,1)$,

$$\overline{\lim_{s\to\infty}}\, s^{3/2}\Big\{E\{(W_1-p_1)^i(W_2-p_2)^{3-i}\}^2\Big\}^{1/2} \leqslant \{d_i(p_1,p_2)\}^{1/4}. \tag{10}$$

where $d_i(p_1,p_2) = \{p_1(1-p_1)\}^{2i}\{p_2(1-p_2)\}^{2(3-i)}\mu_{4i}\mu_{4(3-i)}$.

Either $f_{i,3-i}^{'''} = 0$ or $f_{i,3-i}^{'''} \propto x_1^a/x_2^b$ or $1/x_2^b$, where a and b are natural numbers. If $f_{i,3-i}^{'''} \propto x_1^a/x_2^b$, then, since $\xi_1^{2a}/\xi_2^{2b} \leqslant (p_1^{2a}+W_1^{2a})(1/p_2^{2b}+1/W_2^{2b})$, by the Cauchy-Schwarz inequality and the Minkowski inequality,

$$\begin{aligned} E(\xi_1^{2a}/\xi_2^{2b}) &\leqslant E\{(p_1^{2a}+W_1^{2a})(1/p_2^{2b}+1/W_2^{2b})\} \\ &\leqslant \{E(p_1^{2a}+W_1^{2a})^2\}^{1/2}\cdot\{E(1/p_2^{2b}+1/W_2^{2b})^2\}^{1/2} \\ &\leqslant \left\{p_1^{2a}+\sqrt{E(W_1^{4a})}\right\}\left\{1/p_2^{2b}+\sqrt{E(1/W_2^{4b})}\right\}. \end{aligned}$$

By the Jensen inequality and the Minkowski inequality, one has

$$p_1^{4a} = (EW_1)^{4a} \leqslant EW_1^{4a} = E(W_1-p_1+p_1)^{4a} \leqslant \left[p_1 + \{E(W_1-p_1)^{4a}\}^{1/(4a)}\right]^{4a}.$$

Note that $\lim_{s\to\infty} E(W_1^{4a}) = p_1^{4a}$ since

$$\lim_{s\to\infty}[s/\{p_1(1-p_1)\}]^{2a}E(W_1-p_1)^{4a} = \lim_{s\to\infty} EZ_{1,s}^{4a} = \mu_{4a}.$$

Let $d(u;s,p) = \binom{s}{u} p^u (1-p)^{s-u}$ and write

$$\begin{aligned}\frac{E(1/W_2^{4b})}{2^{4b}} &= \sum_{u=4b-2}^{s} \frac{s^{4b}}{\{u+(u+2)\}^{4b}} d(u;s,p_2) + \sum_{u=0}^{4b-3} \frac{s^{4b}}{\{u+(u+2)\}^{4b}} d(u;s,p_2) \\ &< \frac{s^{4b}}{\prod_{i=1}^{4b}(s+i)} \sum_{u=4b-2}^{s} \prod_{i=1}^{4b} \frac{s+i}{u+i} d(u;s,p_2) + s^{4b} \sum_{u=0}^{4b-3} d(u;s,p_2) \\ &< \frac{1}{p_2^{4b}} \left\{ 1 - \sum_{u=0}^{8b-3} d(u;s+4b,p_2) \right\} + s^{4b} \sum_{u=0}^{4b-3} d(u;s,p_2).\end{aligned}$$

Since $\overline{\lim}_{s\to\infty} E(1/W_2^{4b}) \leqslant (2/p_2)^{4b}$,

$$\overline{\lim_{s\to\infty}} \{E(\xi_1^{2a}/\xi_2^{2b})\}^{1/2} \leqslant \{2(1+2^{2b})p_1^{2a}/p_2^{2b}\}^{1/2}.$$

If $f_{i,3-i}''' \propto 1/x_2^b$, then $\overline{\lim}_{s\to\infty}\{E(1/\xi_2^{2b})\}^{1/2} \leqslant \{(1+2^{2b})/p_2^{2b}\}^{1/2}$. To summarize, there exists $c_i(p_1,p_2)$ always being either zero or positive such that

$$\overline{\lim_{s\to\infty}} E\left|f_{i,3-i}'''(\xi_1,\xi_2)\right|^2 \leqslant c_i(p_1,p_2), \quad i=0,1,2,3. \tag{11}$$

From (8), (9), (10) and (11), conclude that

$$\begin{aligned}\overline{\lim_{s\to\infty}} s^{3/2} \left| E\left\{ f_{i,3-i}'''(\xi_1,\xi_2)(W_1-p_1)^i (W_2-p_2)^{3-i} \right\} \right| &\leqslant c_i^{1/2}(p_1,p_2) d_i^{1/4}(p_1,p_2), \\ \overline{\lim_{s\to\infty}} s^{3/2}|A_2| &\leqslant \sum_{i=0}^{3} \frac{1}{3!} \binom{3}{i} c_i^{1/2}(p_1,p_2) d_i^{1/4}(p_1,p_2).\end{aligned}$$