Dealing with those pesky zeros in ecological models: minding your ps and qs

Table 1. Notation for the composite Bernoulli-Dirichlet model.

n	number of groups
i	group index with $1 \le i \le n$
p_i	nonzero probability of being in group i, based on an ecologial model
$\bar{p}$	mean value $(1/n)$ of $p_1, \ldots, p_n$
q_i	Bernoulli probability of 0 in group i
x_i	Bernoulli random variable (0 or 1) drawn with probability q_i of value 0
p_i'	adjusted proportions p_i, conditioned on x_i
y_i	random proportions from the Dirichlet distribution with parameters $p_1, \ldots, p_n, N$
y_i'	random proportions from the composite Bernoulli-Dirichlet model
$\mathbf{v}$	generic vector notation $\mathbf{v} = (v_1, \ldots, v_n)$; e.g., $\mathbf{p}, \mathbf{p}', \mathbf{q}, \mathbf{x}, \mathbf{y}, \mathbf{y}'$
$\boldsymbol{\theta}$	parameter vector for an ecological model of nonzero probabilities $\mathbf{p}(\boldsymbol{\theta})$
N	effective sample size for a Dirichlet distribution
$a, b, \tilde{q}$	parameters in the relationship between p_i and q_i, where $b > 0$ and $0 < \tilde{q} < 1$
$\boldsymbol{\phi}$	complete model vector $\boldsymbol{\phi} = (\boldsymbol{\theta}, \tilde{q}, b, N)$
$\mathcal{B}(x \vert p)$	Bernoulli distribution, where $x = 0$ or 1 with probability $1 - p$ or p, respectively
$\mathcal{D}(\mathbf{y} \vert \mathbf{p}, N)$	Dirichlet distribution for vector $\mathbf{y}$ of proportions that sum to 1, given the proportion vector $\mathbf{p}$ and the effective sample size N

Table 2. Link between p_i and q_i for $i = 1, \ldots, n$, based on parameters $\tilde{q}$ and b, where $0 < \tilde{q} < 1$ and $b > 0$ and a is computed with (T2.5).

$$\operatorname{logit}(p) \equiv \log\left(\frac{p}{1-p}\right) \tag{T2.1}$$

$$\operatorname{logit}(q_i) = a - b \operatorname{logit}(p_i) \tag{T2.2}$$

$$\bar{p}(\boldsymbol{\theta}) = \frac{1}{n} \tag{T2.3}$$

$$\operatorname{logit}(q_i) = \operatorname{logit}(\tilde{q}) - b\left[\operatorname{logit}(p_i) - \operatorname{logit}(\bar{p})\right] \tag{T2.4}$$

$$a(\boldsymbol{\theta}, \tilde{q}, b) = \operatorname{logit}(\tilde{q}) + b \operatorname{logit}(\bar{p}(\boldsymbol{\theta})) \tag{T2.5}$$

$$q_i(\boldsymbol{\theta}, \tilde{q}, b) = \frac{e^{a - b \log\left(\frac{p_i}{1-p_i}\right)}}{1 + e^{a - b \log\left(\frac{p_i}{1-p_i}\right)}} \tag{T2.6}$$

Table 3. Steps for simulating $\mathbf{y}'(\boldsymbol{\phi})$ from $\mathbf{p}(\boldsymbol{\theta})$ and $\mathbf{q}(\boldsymbol{\theta}, \tilde{q}, b)$, along with the corresponding probability distributions.

$$x_i \sim \mathcal{B}(1 - q_i) \tag{T3.1}$$

$$p'_i = \frac{x_i p_i}{\sum_{i=1}^{n}(x_i p_i)} \tag{T3.2}$$

$$\mathbf{y}'|_{\mathbf{p}'>0} \sim \mathcal{D}(\mathbf{p}'|_{\mathbf{p}'>0}, N) \tag{T3.3}$$

$$P(\mathbf{x}|\mathbf{q}) = \prod_{i=1}^{n}[q_i^{1-x_i}\,(1 - q_i)^{x_i}] \tag{T3.4}$$

$$P(\mathbf{y}|\mathbf{p}, N) = \frac{\Gamma(N)}{\prod_{i=1}^{n}\Gamma(Np_i)}\prod_{i=1}^{n} y_i^{Np_i - 1} \tag{T3.5}$$

$$P(\mathbf{y}'|\mathbf{p}', N) = \frac{\Gamma(N)}{\prod_{p'_i>0}\Gamma(Np'_i)}\prod_{p'_i>0} y_i'^{\,Np'_i - 1} \tag{T3.6}$$

Table 4. Inference functions for estimating $\boldsymbol{\phi}$ from observed data $\mathbf{y}'$, where L denotes the likelihood and ℓ the negative log likelihood.

$$x_i = \begin{cases} 0 \text{ if } y'_i = 0 \\ 1 \text{ if } y'_i > 0 \end{cases} \tag{T4.1}$$

$$p'_i(\boldsymbol{\theta}, \mathbf{x}) = \frac{x_i p_i(\boldsymbol{\theta})}{\sum_{i=1}^{n}[x_i p_i(\boldsymbol{\theta})]} \tag{T4.2}$$

$$L(\boldsymbol{\phi}|\mathbf{y}') = P[\mathbf{y}'|\mathbf{p}'(\boldsymbol{\theta}, \mathbf{x}), N]\,P[\mathbf{x}|\mathbf{q}(\boldsymbol{\theta}, \tilde{q}, b)] \tag{T4.3}$$

$$\begin{aligned} \ell(\boldsymbol{\phi}) = &\sum_{p'_i>0}[\log\Gamma(Np'_i) - (Np'_i - 1)\log y'_i] \\ &- \sum_{i=1}^{n}[(1 - x_i)\log q_i + x_i\log(1 - q_i)] - \log\Gamma(N) \end{aligned} \tag{T4.4}$$

Table 5. Model notation for simulating proportions p_i in a stable fish age distribution.

i	age class index ($i = 1, \ldots, n$)
n	maximum age class
Z	total mortality from natural causes and fishing ($Z = M + F$)
S_i	survival from recruitment (age class 1) to age class i
β_i	fishery selectivity on age class i ($0 < \beta_i \le 1$)
α	selectivity parameter ($\alpha > 0$)
A	age class with full selectivity, where $\beta_i = 1$ for $A \le i \le n$
R_i	historical recruitment for age class i
B	age class with anomolous recruitment
$\boldsymbol{\delta}$	fixed design vector $\boldsymbol{\delta} = (n, A)$
$\boldsymbol{\theta}$	vector $\boldsymbol{\theta} = (Z, \beta_1, \alpha)$ of potentially estimable parameters
$p_i(\boldsymbol{\theta}\|\boldsymbol{\delta})$	simulated proportion in age class i
$\boldsymbol{\phi}$	full parameter vector for the Bernoulli-Dirichlet model, in this case $\boldsymbol{\phi} = (Z, \beta_1, \alpha, \tilde{q}, b, N)$

Table 6. Model equations for computing $\mathbf{p}(\boldsymbol{\theta})$, given the design vector $\boldsymbol{\delta}$.

$$S_i = \begin{cases} e^{-Z(i-1)}; & i = 1, \ldots, n-1 \\ \dfrac{e^{-Z(i-1)}}{1 - e^{-Z}}; & i = n \end{cases} \tag{T6.1}$$

$$\beta_i(\beta_1, \alpha) = \begin{cases} 1 - (1 - \beta_1)\left(\dfrac{A - i}{A - 1}\right)^{\alpha}; & i = 1, \ldots, A-1 \\ 1; & i = A, \ldots, n \end{cases} \tag{T6.2}$$

$$R_i = 1 \tag{T6.3}$$

$$p_i(\boldsymbol{\theta}) = \frac{S_i \beta_i R_i}{\sum_{i=1}^{n} S_i \beta_i R_i}; \qquad i = 1, \ldots, n \tag{T6.4}$$

GUI: Entries for $\boldsymbol{\delta}$ and $\boldsymbol{\phi}$. Buttons for simulation, estimation, and MCMC. Probably use a notebook with a second page for graphics output options. Source my example `pq.r`.

Simulation: Calculate $\mathbf{p}$ from Table 6, based on $\boldsymbol{\delta}$ and $\boldsymbol{\theta}$. Then get $\bar{p}$, a and $\mathbf{q}$ from (T2.3), (T2.5) and (T2.6). Generate $\mathbf{x}$ from (T3.1) using `rbinom` with probability $1 - q_i$. (This corresponds to the distribution (T3.4).) Use (T3.2) to compute $\mathbf{p}'$. Finally, use (T3.3) to produce the Dirichlet sample $\mathbf{y}'$. (This corresponds to the distribution (T3.6).) Define `rdirich` using `rgamma` as in the past. Write $\mathbf{y}'$ to a file for use by ADMB in the estimation step.

Estimation: Start with $\mathbf{y}'$ from the simulation. Get $\mathbf{x}$ from (T4.1). Use (T4.4) to calculate the negative log likelihood ℓ needed by ADMB. Always compute `log gamma`, not `gamma` directly (`lgamma` in R and `gammaln` in ADMB).