N$_c$混合占据模型

Huu-Dinh Huynh[1],Wen-Han Hwang[2]

[1]国立中兴大学统计学研究所，台湾台中[2]国立清华大学统计学研究所，台湾新竹wenhan@stat.nthu.edu.tw

arXiv:2304.02851v1 [stat.ME] 6 Apr 2023

摘要

本文提出一类用于检测/未检测数据的占据模型，以放宽 N–混合模型的封闭性假设。我们引入一个取值范围为 0 到 1 的群落参数 c，用于描述在多次访问中保持固定的个体所占的比例。因此，当 c 等于 1 时，该模型退化为 N–混合模型；研究表明，当封闭性假设未得到完全满足时，这一简化模型会高估丰度。此外，通过纳入零膨胀成分，所提出的模型能够衔接标准占据模型（$c = 0$）与零膨胀 N–混合模型（$c = 1$）。随后，我们研究了当 c 从 0 变化到 1 时这两种极端模型的估计量表现。一个有趣的发现是，当 c 趋近于 0 时，零膨胀 N–混合模型能够一致地估计零膨胀概率（占据率）；但当 $c > 0$ 时，其偏差可能为正、为负或为零，具体取决于其他参数。我们还通过模拟研究和数据分析展示了这些结果。

关键词：群落参数；N–混合模型；零膨胀模型

1 引言

在生态保护与管理中，尽管探测并不完善，估算特定区域内物种的占据率和丰度仍是一项关键问题。实际上，物种可能存在于调查地点，却因调查方法或调查时机的局限而未被探测到。地点抽样调查中的零计数，可能是因为物种不存在，也可能是探测错误所致。若不处理这一问题，占据率就会被低估。基于多次访问的探测/未探测（出现、有无）数据，地点占据模型（MacKenzie 等，2002）通过考虑探测错误来估算出现率。这些模型可用于时间或空间重复调查，且无需识别或标记个体，因此具有成本效益。它们广泛应用于物种分布建模，并已发展出多种扩展模型，例如多季节开放模型、多物种模型、动态模型和时空模型（MacKenzie 等，2017；Hogg 等，2021；MacKenzie 等，2009；Johnson 等，2013）。本研究聚焦于单物种、单季节占据模型。

在标准占据模型（MacKenzie 等，2002）中，检测概率与物种丰度相互混杂，无法彼此区分。为克服这一局限，Royle 和 Nichols（2003）；Royle（2004）提出了 N–混合占据模型，使检测概率与物种丰度得以分离。这些模型能够通过多次调查的出现数据或计数数据估算物种丰度。N–混合模型在文献中受到了广泛关注（Haines，2016a,b；Joseph 等，2009；Gomez 等，2018），因为它们能够像捕获—再捕获模型一样估算种群规模，却无需捕获和标记个体。然而，这些模型的表现取决于所作的假设（Dennis 等，2015；Link 等，2018；Barker 等，2018）。因此，我们的目标是通过放宽闭合模型假设来改进这些模型；即使在单季调查中，这一假设也经常被违反（Kendall 等，2013；Otto 等，2013）。

封闭性假设指出，在多次访问期间，一个地点的个体数量保持不变。调查通常通过时间重复调查或空间重复调查进行，有时还会使用多个探测器，或结合这些方法（MacKenzie 等，2017 年；Kendall 和 White，2009 年）。时间重复调查涉及在不同时间对相同地点进行调查，而空间重复调查则涉及在单个地点内的大范围区域中随机选择抽样单元。虽然对于在短时间内于相同地点开展的调查而言，这一封闭性假设是合理的，但它可能并不适用于高度移动的物种（Hayes 和 Monfils，2015 年）。我们还注意到，Royle（2004 年）的研究是时间重复调查的一个例子，尽管论文标题提到了空间重复调查，这是为了说明地点的分布。

事实表明，当封闭性假设被违反时，N–混合占据率模型的推断会产生有偏的点估计和不正确的区间估计。尽管少数研究已强调了这些影响（D´enes 等人，2015年；Duarte 等人，2018年；Dail 和Madsen，2011年；Ke 等人，2022年），但大多数证据都来自模拟研究。据我们所知，目前缺乏解释 N–混合模型下估计偏差行为的理论结果。然而，我们最近在所提出的 N_c–混合模型下提供了理论证据，该模型是 N–混合模型的扩展。需要注意的是，Dail 和 Madsen（2011年）还提出了一类广义N–混合模型，该模型允许物种种群迁入和迁出，并估计逐年迁入率和迁出率，为保护管理者提供了有价值的洞见。然而，该模型是一个多季节开放

与我们的扩展方向不同的种群模型。

N_c–混合模型旨在建立一个框架，将时间重复调查和空间重复调查统一起来。例如，考虑在某个地点进行的一项三次访问调查，其中 N_1、 N_2 和 N_3 分别表示每次访问期间可观测个体的数量。假设这些变量是同分布的随机变量。对于时间重复调查，由于调查在短时间内进行，三个 N_j变量可视为相等，并满足 N–混合模型的封闭性假设。相比之下，对于空间重复调查， N_j 变量则视为相互独立。为兼顾这两种情形，我们将 N_j 变量分解为两个部分： $N_j = K + M_j$ 对于$j = 1$，2，3，其中 K 表示三次访问调查中的共有个体数量，而 M_j 表示非共有个体数量。我们假设 K 和M_j 相互独立，且对于某个 $0 \leq c \leq 1$, $E(K) = cE(N_j)$。参数 c称为群落参数，表示在三次访问期间作为常驻个体且保持不变的个体所占的比例。该参数还使我们可以轻松地在 $c = 0$时令 K 退化为 0，或在 $c = 1$时令 M_j 退化为 0。图 1 展示了这一分解。

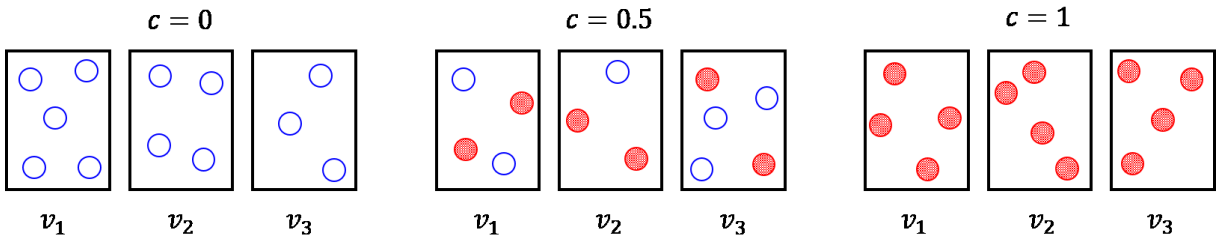


图1：N_c–混合模型的示意图。在对一个地点进行三次访问的调查中，有些个体在三次访问期间都存在于该地点（分别表示为 v_1、 v_2和 v_3）。该图展示了对应于群落参数 $c = 0$（左）、$c = 0.5$ （中）和 $c = 1$ （右）的情形。彩色圆圈表示在三次访问期间为该地点的常驻个体（或者等价地说，是固定的个体）。对于 $c = 0$，个体数量可能随访问次数而变化，模拟空间重复调查的情景。对于 $c = 1$，个体数量保持不变，且满足N–混合模型的封闭性假设。对于 $c = 0.5$，约有一半的个体在调查期间为常驻个体。

在 N_c–混合模型下，我们能够证明，如果将群落参数 c 错误地指定为 1（即改用N–混合模型），则平均丰度将被高估。此外，我们的结果表明，偏差会随着群落参数 c的减小而增大，并在 c 趋近于 0 时达到无穷大。

我们提出了对 N_c–混合模型的扩展，引入了一个零膨胀成分。当 $c = 0$（MacKenzie 等，2002）时，该模型可与标准占据模型衔接；当 $c = 1$ （Haines, 2016a）时，则可与零膨胀 N–混合模型衔接。随后，我们研究了这两个极端模型的估计量在群落参数 c从 0 变化到 1 时的表现。研究结果表明，标准占据模型会低估零膨胀概率（占据率），且偏差会随着群落参数 c 的增加而增大。然而，一个有趣的发现是，当 c 趋近于 0 时，零膨胀 N–混合模型能够一致地估计占据率，但偏差的正负取决于其他参数。

本文的组织结构如下：第 2 节中，我们构建 N_c–混合模型并介绍估计方法。第 3 节中，我们考虑零膨胀 N_c–混合模型。第 4 节包括用于评估估计量性能的模拟研究。两组真实数据

示例见第 5 节，讨论见第 6 节。网络附录 A 包含所有命题的证明。

2 N_c–混合模型

2.1 符号与估计

考虑一项多次访问抽样调查，包含 n 个地点和 T 次访问。设 N_{ij} 为第 j 次访问期间地点 i 的个体数。根据引言中的动机（见图 1），我们将 N_{ij} 分解为两个相互独立的泊松分布随机变量： K_i 和 M_{ij}。 K_i 的期望值为 $c\mu$ ，而 M_{ij} 的期望值为 $(1-c)\mu$，其中 $0 \leq c \leq 1$ 和 $\mu > 0$ 为常数参数。因此，每个地点的个体数 N_{ij}服从均值为 μ的同一泊松分布，该均值表示各地点（每次访问）的丰度参数。然而，同一地点多次访问所得的个体数 N_{ij}、 $j = 1$、 $\cdots$、 T并非相互独立，因为它们共享一个共同变量 K_i。群落参数 c 刻画访问次数之间的依赖程度，因为它也表示 N_{ij}每次访问之间的相关性。注意，当 $c = 1$ 和 $c = 0$时， M_{ij}和 K_i 分别退化为 0。

为对数据观测过程建模，我们假设每个个体都可以被独立探测，探测概率为 r。如果在地点 i 第 j 次调查期间探测到该物种，则令 $Y_{ij} = 1$，否则令 $Y_{ij} = 0$。我们将地点 i 第 j 次调查的探测概率记为 $P(Y_{ij} = 1) = p_{ij}$，那么当已知 N_{ij}时， $p_{ij} = 1-(1-r)^{N_{ij}}$。这构成了一个 N_c–混合模型。

显然， Y_{ij}、 $j = 1$、 $\cdots$、 T是可交换变量，不过除非满足 $c = 0$，否则它们并不独立。令 $Y_i = \sum_{j=1}^{T} Y_{ij}$ 表示地点 i的出现频率。在网络附录 A 中，我们证明 N_c–混合模型下 Y_i 的概率函数为

$$f(y_i;\boldsymbol{\theta}) = \binom{T}{y_i} \exp(-d\mu rT - c\mu) \sum_{k=0}^{y_i} \binom{y_i}{k} (-1)^k \exp\left\{c\mu(1-r)^{T-y_i+k} + d\mu r(y_i-k)\right\}, \quad (1)$$

其中 $\boldsymbol{\theta} = (\mu, r, c)$ 是模型参数向量，且 $d = 1 - c$。方程（1）也称为 N_c–混合模型。

当群落参数 $c = 1$ 时，方程 (1) 中的模型简化为

$$f(y_i;\boldsymbol{\theta}_1) = \binom{T}{y_i} \exp(-\mu) \sum_{k=0}^{y_i} \binom{y_i}{k} (-1)^k \exp\left\{\mu(1-r)^{T-y_i+k}\right\}, \quad (2)$$

其中 $\boldsymbol{\theta}_1 = \boldsymbol{\theta}$ 与 $c = 1$。这是 N–混合模型的概率函数（Royle 和 Nichols，2003），但显式形式（2）最早由 Haines（2016a）给出。

同样地，当群落参数 $c = 0$ 时，模型（1）简化为

$$f(y_i;\boldsymbol{\theta}_0) = \binom{T}{y_i} \{\exp(-\mu r)\}^{T-y_i} \{1-\exp(-\mu r)\}^{y_i}, \quad (3)$$

其中， $\boldsymbol{\theta}_0 = \boldsymbol{\theta}$ ，且 $c = 0$。简化模型(3)是二项分布，记作$Y_i \sim \mathrm{Bino}(T, p)$，其中 $p = 1 - \exp(-\mu r)$。我们还注意到，二元变量 Y_{ij} 独立同分布，其均值为 $E\{1-(1-r)^{N_{ij}}\}$，当 $c = 0$时。利用泊松矩生成函数进行计算可得， $E\{1-(1-r)^{N_{ij}}\} =$

$1-\exp(-\mu r)=p$，这会得出与 (3) 相同的结果。然而，(3) 中的模型参数 μ和 r 无法分离，因此该模型不可识别，只有它们的乘积$\mu \times r$ 可识别。

当群落参数 $0<c<1$时，N$_c$–混合模型 (1) 对于 $T \geq 3$是可识别的。参数向量 $\boldsymbol{\theta}$ 的对数似然函数为 $\ell(\boldsymbol{\theta})=\sum_{i=1}^{n} \log\{f(y_i;\boldsymbol{\theta})\}$，最大似然估计很容易求得。为便于后续使用，可以在多项式模型框架下表示似然函数。设$m_j=\sum_{i=1}^{n} \mathrm{I}(y_i=j)$，其中 $j=0,\cdots,T$，I(·) 为指示函数。这些统计量 m_j可视为具有单元概率 $f(j;\boldsymbol{\theta})$ 的多项式模型的结果。因此，我们可以写出 $\ell(\boldsymbol{\theta})=\sum_{j=0}^{T} m_j \log\{f(j;\boldsymbol{\theta})\}$，而 $\boldsymbol{\theta}$的得分函数为

$$S(\boldsymbol{\theta})=\sum_{j=0}^{T} \frac{\partial f(j;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \left\{ \frac{m_j-nf(j;\boldsymbol{\theta})}{f(j;\boldsymbol{\theta})} \right\}. \tag{4}$$

2.2 占据率

Royle 和 Nichols（2003）将占据率 ψ 定义为 N–混合模型下的一个派生参数。具体而言，按照我们的记号，$\psi=P(K_i>0)=1-\exp(-\mu)$。在 N$_c$–混合框架下，每次访问的占据率可定义为$P(N_{ij}>0)=1-\exp(-\mu)$。或者，也可以将该比率定义为 $1-P(N_{ij}=0,\ \forall\ 1\leq j\leq T)$；若某些 $N_{ij}>0$，则认为该地点 i 被该物种占据。由此，占据率为 $1-\exp\{-\mu-(T-1)d\mu\}$，它取决于访问次数 T，并在 T 趋于无穷时收敛于 1。因此，这两种定义都不同于当前的地点占据概念。问题在于，由于在 N$_c$–混合模型中，M_{ij}, $j=1,\ldots,T$都是随机变量，地点 i的个体数量可能会随访问次数而变化。因此，我们不直接定义占据状态，而是希望引入一个零膨胀参数，以确定 N$_c$–混合模型下的地点占据状态。第 3 节将讨论这一扩展。

2.3 N–混合模型估计量的行为

在本小节中，我们考察当多次访问期间某个地点的个体数量不是固定常数（即 $c<1$）时，μ 和 r 的 N–混合模型估计量的表现。

首先，我们考察群落参数 c 已知且 $T=2$（双重访问）的情形。在此情形下，我们使用记号 $\tilde{\mu}_c$ 和 $\tilde{r}_c$ 分别表示 μ 和 r的（受限）最大似然估计量（最大似然估计）。如网络附录A 的网络引理1 所示，当 $T=2$时，最大似然估计量具有闭式解，分别为 $\tilde{\mu}_c=cz_1^2/(2z_1+z_2)$和 $\tilde{r}_c=(2z_1+z_2)/(cz_1)$，其中 $z_1=\log\{2n/(2m_0+m_1)\}$ 以及 $z_2=\log(m_0/n)$。这些闭式解使我们能够轻松确定N–混合模型的最大似然估计量 $\tilde{\mu}_1$ 和 $\tilde{r}_1$的极限，如下述命题所示。

命题1. 在采用双次访问调查的 N$_c$–混合模型下，随着地点数量趋于无穷大，估计量 $\tilde{\mu}_1$ 和 $\tilde{r}_1$ 分别以概率为 1 收敛于 μ/c 和 cr，对所有 $0<c\leq 1$ 均成立。

命题1揭示了群落参数小于1时N–混合模型估计量的行为。结果表明，当 $c=$时，最大似然估计是一致的。

1，这是最大似然方法的一个常见特性。此外，结果表明，N–mixture 模型高估了丰度参数 μ，且偏差会随着群落参数 c 的降低而增大。相比之下，N–mixture模型在检测概率 r方面呈现出相反的偏差行为。尽管存在这些偏差，估计量 $\tilde{\mu}_1 \times \tilde{r}_1$仍能在 N_c–mixture 模型的框架下持续估计 $\mu \times r$，这是值得注意的。

我们最初认为，命题1的结果适用于访问次数超过两次的调查（$T > 2$），但进一步研究发现，这一说法仅部分正确。正确之处在于，N–混合模型存在一些矩估计量，其表现与命题1中所述的表现相似。为证明这一点，我们推导了N–混合模型下 Y_i 的矩，并将所得估计量定义为 $\tilde{\mu}_{1\mathrm{M}}$ 和$\tilde{r}_{1\mathrm{M}}$。以下命题总结了这一分析的结果。

命题2。N–混合模型的矩估计量 $\tilde{\mu}_{1\mathrm{M}}$ 和 $\tilde{r}_{1\mathrm{M}}$通过求解以下方程获得

$$\bar{Y} = Tp \quad and \quad \bar{Y}^2 = Tp + T(T-1)\{2p - 1 + (1-p)^{2-r}\},$$

其中，$\bar{Y}$为样本平均值，且$\bar{Y}^2 = \sum_{i=1}^{n} Y_i^2/n$。此外，如果$N_c$–混合模型成立，随着地点数量趋于无穷，估计量 $\tilde{\mu}_{1\mathrm{M}}$ 和 $\tilde{r}_{1\mathrm{M}}$ 分别以概率为1收敛于μ/c 和 cr，对所有 $0 < c \leq 1$均成立。

对于多次访问调查，我们的模拟研究（第 4 节）发现，当 $0 < c < 1$ 时，N–混合模型的最大似然估计 $\tilde{\mu}_1$ 的极限呈现出与 μ/c 相似的模式。然而，两者之间存在差异，而且我们只能确定群落参数 c 接近 0 时的行为。

命题 3。在 c 趋近于零的 N_c–混合模型下，随着地点数量 n 趋于无穷大，估计量 $\tilde{\mu}_1$ 和 $\tilde{r}_1$ 收敛，使得 $\tilde{\mu}_1 \to \infty$、$\tilde{r}_1 \to 0$，且 $\tilde{\mu}_1\tilde{r}_1 \to \mu r$ 的概率为 1。

根据所呈现的结果，我们认为，当 $c < 1$ 时，N–混合模型的最大似然估计倾向于高估参数 μ 并低估参数 r，且随着 c 偏离 1 的程度增大，偏差也随之增大。然而，在 N_c–混合模型框架下，估计量 $\tilde{\mu}_1 \times \tilde{r}_1$ 仍可能在某些 c 取值范围内一致地估计 $\mu \times r$，例如 c 接近 0 时。

3 零膨胀 N_c– 混合模型

为考虑物种占据率，我们通过纳入零膨胀成分对模型(1) 进行扩展。按照MacKenzie等人（2002）的做法，令 ψ 为地点占据概率，则零计数（$Y_i = 0$）的概率为（$1 - \psi$） $+ \psi f$(0; $\boldsymbol{\theta}$)。因此，似然函数增加了一个参数 ψ，从而可以写为

$$L(\boldsymbol{\theta}, \psi) = \prod_{i=1}^{n} \{(1-\psi)\mathbf{I}(y_i = 0) + \psi f(y_i; \boldsymbol{\theta})\}. \tag{5}$$

我们将模型（5）称为零膨胀 N_c–混合（ZIN$_c$）模型。

当 $c = 1$ 时，模型（5）变为零膨胀 N–混合模型（ZIN模型），如 Haines（2016a）先前所述。在 $c = 0$ 的情况下，它简化为零膨胀二项分布

（ZIB）模型，其检测概率为 $p = 1 - \exp(-\mu r)$，也称为第一个占据率模型（MacKenzie 等，2002）。因此，ZIN$_c$ 模型将 ZIB 和 ZIN 模型统一到同一个框架中，其中 $c = 0$ 和 $c = 1$ 的特殊情况分别对应 ZIB 和 ZIN。

3.1 估计

当群落参数 $0 < c < 1$时，ZIN$_c$ 模型对于 $T \geq 4$ 是可识别的。通过定义 $p_0(\boldsymbol{\theta}, \psi) = (1 - \psi) + \psi f(0; \boldsymbol{\theta})$ 和 $f(+; \boldsymbol{\theta}) = 1 - f(0; \boldsymbol{\theta})$，似然函数可以写为

$$L(\boldsymbol{\theta}, \psi) = L_0(\boldsymbol{\theta}, \psi) \times L_1(\boldsymbol{\theta}),$$

其中 $L_0(\boldsymbol{\theta}, \psi) = \{p_0(\boldsymbol{\theta}, \psi)\}^{m_0}\{1 - p_0(\boldsymbol{\theta}, \psi)\}^{n-m_0}$ 以及 $L_1(\boldsymbol{\theta}) = \prod_{j=1}^{T} \{f(j; \boldsymbol{\theta})/f(+; \boldsymbol{\theta})\}^{m_j}$。注意，$L_0$ 反映了I($Y_i > 0$) 的概率函数，而 L_1 是基于 $Y_i > 0$的条件似然。由于条件似然函数独立于占据率概率 ψ，因此可以在不受 ψ 混淆的情况下估计 $\boldsymbol{\theta}$（Karavarsamis and Huggins, 2020）。具体而言，我们可以通过求解 $L_1(\boldsymbol{\theta})$ 的得分函数来找到 $\boldsymbol{\theta}$的条件最大似然估计，记为 $\hat{\boldsymbol{\theta}}$。此外，通过对剖面似然 $L_0(\boldsymbol{\theta}, \psi)$ 求最大似然估计，我们可以找到 $\hat{\psi} = (n - m_0)/\{nf(+; \hat{\boldsymbol{\theta}})\}$。如网络附录 A 所示，网络引理 2 确认，由该方法得到的估计量 $\hat{\boldsymbol{\theta}}$ 和 $\hat{\psi}$ 也是基于（5）的通常最大似然估计。$\hat{\boldsymbol{\theta}}$ 和 $\hat{\psi}$ 的渐近方差可以按通常方式推导。

在零膨胀类型模型中，主要关注的是占据率而非概率的估计。设在已知群落参数 c 的情况下，$\hat{\psi}_c$ 是 ZIN$_c$ 模型中 ψ 的最大似然估计。接下来，我们考察 $\hat{\psi}_0$ 和 $\hat{\psi}_1$的性质，它们分别对应 ZIB 和 ZIN 模型的占域估计量。

3.2 $\hat{\psi}_0$ 和 $\hat{\psi}_1$ 的行为

与N$_c$–混合模型一样，ZIN$_c$ 模型在 $c = 0$时，μ 和 r也存在可辨识性问题。因此，在ZIB模型下，只能估计参数 p（即 $1 - \exp(-\mu r)$）和占域概率 ψ。实际上，为了拟合ZIB模型，通常将 $r = 1$ 设为以估计 $\mu \times r$，所得的丰度估计量为（$\tilde{\mu}_0$）。

命题4。在 ZIN$_c$ 模型中，当 $c > 0$时，ZIB占据率估计量 $\hat{\psi}_0$ 随着 n 增大至无穷大，以概率为1 表现出对 ψ 的低估。该低估的线性近似可表示为 $\hat{\psi}_0 \approx \frac{p}{p+\Delta}\psi$，其中

$$\Delta = \frac{\{1 - (1-p)^T\}\left\{f(0;\boldsymbol{\theta}) - (1-p)^T\right\}}{\left[\frac{1}{p}\{1 - (1-p)^T\} - T(1-p)^{T-1}\right]\{1 - f(0;\boldsymbol{\theta})\}}, \tag{6}$$

当 c 增加时，Δ 也随之增加；当 $c = 0$ 时，$\Delta = 0$。

我们注意到，随着 T 或 $\mu \times r$ 增加，Δ 的值会降至零。当访问次数增多或物种丰度较高时，观测占据率会趋近于 ψ，因此这一结果是预期之中的。线性近似偏差合理地呈现了 $\hat{\psi}_0$ 在各个方面的趋势，表明随着 c 偏离0（或随着访问间相关性增加），低估会变得更加明显。

注释 1. 作为命题 4 的直接推论，我们还可以看出，相应的丰度估计量 $\tilde{\mu}_0$ 会随着 c 的增加而增加，并且在 $c=0$ 处的 $\tilde{\mu}_0$ 是物种丰度与检测概率乘积（$\mu \times r$）的一致估计量。

当群落参数 c小于 1 时，ZIN模型的估计量往往会产生偏差。然而，这些偏差的变化规律较为复杂。例如，随着 c减小，占域估计量 ψ_1 的偏差并非单调变化。尽管如此，当 $c \approx 0$ 时，ZIN 模型的估计量（ ψ_1 除外）表现与 N– 混合模型的估计量相似，如命题 3 所示。有趣的是， ψ_1 可以在 $c=0$下一致估计 ψ，如下一条命题所示。为清楚起见，若无歧义，本命题中的 $\tilde{\mu}_1$ 和 $\tilde{r}_1$ 也指 ZIN模型的最大似然估计（或 ZIN$_c$模型的受限最大似然估计）。

命题 5. 在 N$_c$–混合模型下，当 c 趋近于零时，随着地点数量 n 增加，估计量 $\tilde{\mu}_1$ 和 $\tilde{r}_1$ 收敛，使得 $\tilde{\mu}_1 \to \infty$、 $\tilde{r}_1 \to 0$，并且 $\tilde{\mu}_1 \tilde{r}_1 \to \mu \tilde{r}$ 以概率为 1。此外，当 $c=0$ 时， ψ_1 是 ψ 的一致估计量。

估计量 ψ_1 在ZIN（或ZIN$_c$ 且 $c=1$）模型下也是一致估计量。因此，命题5表明，ψ_1 表现得像一座桥梁，其两端（位于 $c=0$ 和1）处于同一水平；然而，目前尚不清楚桥面是否始终位于该水平之上或之下，还是在不同区段有所变化。我们将通过一项模拟研究进一步探究这种行为。

3.3 ZIN 和 ZIB 检验

ZIN或ZIB模型的原假设可以在ZIN$_c$模型的框架下进行检验，因为这等价于检验 $\{c=1\}$或 $\{c=0\}$的取值。一种简单的假设判断方法是使用基于 c估计值的Wald 型置信区间。更正式的方法是使用似然比检验，因为原假设是完整ZIN$_c$ 模型的子模型。然而，需要注意的是，原假设下似然比检验的渐近分布是0与卡方分布的混合，而不是通常的卡方分布（Self和Liang，1987）。在实践中，我们还建议在原假设下生成自助样本，以求得 p–值。这与Dail 和 Madsen（2011）的结论相似。

4 模拟研究

我们进行了模拟，以评估所提出模型和估计量的性能。我们考虑了两种情景：N$_c$–混合模型和 ZIN$_c$ 模型。在第一种情景中，我们分别计算了 N–混合模型和 N$_c$–混合模型的最大似然估计量。在第二种情景中，我们计算了 ZIB、ZIN 和 ZIN$_c$ 模型的最大似然估计量。所有估计值均使用 `R` 软件中的`optim` 函数计算（R Core Team，2022年）。

模拟中，我们将真实参数值设为 $\mu=1$、2、 $r=0.25$、0.5 和 $c=0$、0.05、.....、1。对于两种情景，地点数量分别设为 $n=200$、500、1000，访问次数分别设为 $T=5$、7、10。在第二种情景中，我们还将占域概率设为 $\psi=0.7$。针对每组参数设置，我们生成了 1,000 个数据集，并计算每个数据集的估计值及相应的标准误估计值。

为考虑部分模拟数据集中的离群值，我们报告参数估计值的中位数（Med）、估计标准误的中位数（Med.se），以及经过缩放以与正态分布相匹配的中位绝对偏差（MAD）。此外，我们还报告名义95% Wald 型置信区间的覆盖率（CP）。我们注意到，在某些情况下，用于估计各模型参数的数值方法未能收敛；ZIN模型的未收敛情况更为频繁，而 ZIN$_c$ 模型的未收敛情况较少。不过，在大多数情况下，失败比例很低，因此未予报告。

4.1 模拟研究A：N$_c$–混合模型

$T = 7$ 和 $n = 500$ 的 μ 中位数估计值见图2。Web 附录B中的Web表1–6全面展示了 $c =$ (0.25、0.5、0.75)的模拟结果。

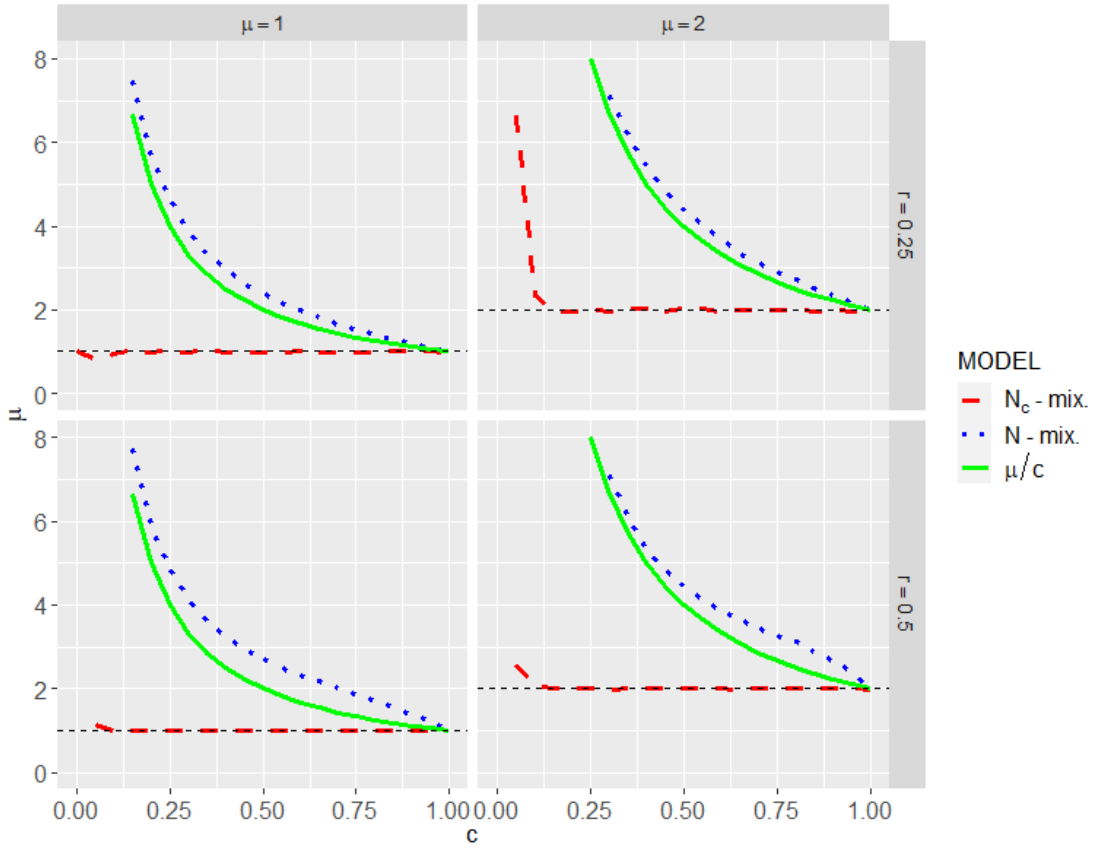


图2：N–混合模型和N$_c$–混合模型中丰度 μ 的中位数估计值（模拟研究A）随群落参数 c 的变化，访问次数$T = 7$ 和地点数 $n = 500$如图所示。子图对应 $\mu = 1$为2、个体探测概率 $r = 0.25$为0.5的四种组合。为清晰起见，已删除取值较高的数据点。

图2展示了N–混合丰度估计量 $\tilde{\mu}_1$ 作为群落参数 c的函数时的表现。该图显示，对于所有小于1的 c 值，该估计量都呈现正偏差。这种偏差呈单调递减趋势，并且只有当 c 等于1时才趋近于真实值。当 c 接近0时，偏差可能很大，但为便于可视化，已从图中移除。此外，该图显示

当 $\tilde{\mu}_1$ 小于 1 时，对于所有 c 的取值，$\tilde{\mu}_1$ 都大于 μ/c，不过尚未明确探讨二者之间的差异。在另一项未报告的模拟研究中，我们计算了 N– 混合模型的矩估计量，发现其模式与参考曲线 μ/c 非常相似。这些模拟结果与命题 1-3 中概述的发现一致。根据 Web 附录 B 的 Web 图 1 中所示的估计量 $\tilde{r}_1$ 的结果，也可以得出相应的结论。N_c– 混合模型的最大似然估计可以一致地估计 μ，但也经常在 $c = 0$ 附近表现出偏差。这主要是因为当 $c \approx 0$ 时，μ 和 r 的参数几乎无法识别。当 μ 较大且 r 较小时，偏差更明显；但随着 n 或 T 增大，偏差会减小；参见 Web 表 6（针对 $n = 1000$ 和 $T = 10$ 的情形）。

从网页表1至6中可以看出，当 c 从0.75变为0.25时，N–混合丰度估计量的相对偏差在40%至400%之间。当 T 或 r 增加时，偏差更为明显；而当 μ 增加时，偏差则有所减轻。另一方面，N_c–混合模型在所有情况下都显示出近乎无偏的估计，唯有 $n = 200$、$T = 5$和$r = 0.25$时例外：当 μ 从2变为1时，相对偏差最高可达7.5%至16%；详见网页表1和4。

就平均绝对偏差（MAD）而言，随着 c增加，两个模型的变异程度均有所下降，其中N–混合模型的下降幅度明显大于N_c–混合模型。具体而言，当 $c = 0.25$时，$\tilde{\mu}_1$ 的MAD可能是$\hat{}\mu$的2至4倍；但当 $c = 0.75$时，前者通常小于后者。由Med.se测得的渐近标准误估计通常与相应MAD的结果一致，因此对于所考察的情形而言具有较好的可靠性。当数据信息充足时，N_c–混合模型的Wald型置信区间估计量表现良好；在 $n \geq 500$ 和 $T \geq 7$时，其结果与名义95%的置信水平较为接近。然而，在某些 r 和 c较小的情况下，覆盖概率（CP）可能低于80%（如网页表1、4和5所示），表明其表现不够理想。当 $c \leq 0.75$时，由于 $\tilde{\mu}_1$ 存在严重偏差问题，N–混合模型估计量的覆盖概率常常达到0。

网络表 1-6 中关于检测概率参数 r 的结果与 μ相似，唯一的区别是 N–混合模型估计量的偏差方向相反。最后发现，在所有情况下，乘积 $\tilde{\mu}_1 \times \tilde{r}_1$的中位数对于估计 $\mu \times r$ 几乎没有偏差。请注意，这一性质仅在范围 $c \approx 0$ 内得到证明（命题 5），但模拟结果表明，c 的范围或许可以扩展。

4.2 模拟研究 B：ZIN_c 模型

图 3 和图 4 展示了参数 μ 和 ψ 在$T = 7$ 和 $n = 500$时的中位数估计结果。关于 $c =$ (0.25、0.5、0.75)的详细模拟结果见网络附录 B 的网络表 7-12。在 ZIB模型中，我们固定了$r = 1$的值。这是因为在这种情况下，参数 $\mu \times r$ 不可分离。

在图3 中，我们可以看到 ZIB估计量 $\tilde{\mu}_0$ 始终低估 μ。事实上，$\tilde{\mu}_0$ 在 c接近0 时大约等于 $\mu \times r$，并且随着 c增加而增加（如备注1 所述）。当 c接近 0 时，ZIN估计量 $\tilde{\mu}_1$ 与 N–混合模型呈现出类似的趋势，但随后迅速下降、缓慢上升，并在 $c = 1$处与 μ 一致。一般而言，$\tilde{\mu}_1$ 主要表现出正偏差，但在某些$0 < c < 1$处偶尔也会表现出负偏差。ZIN_c 估计量$\hat{}\mu$ 通常在c范围的两端都表现出一定偏差，当 r 较小时，在 $c \approx 0$ 处的偏差可能非常大。正如预期的那样，增加 n 和 T 可以

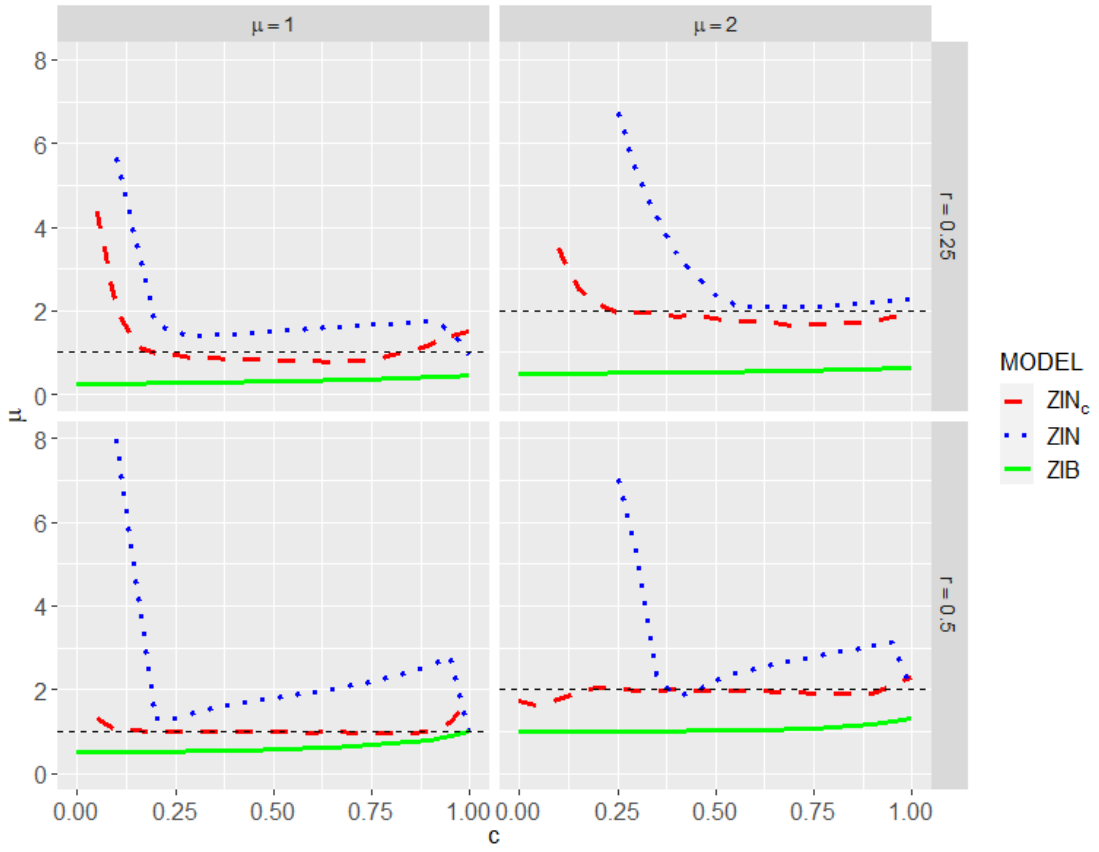


图 3：随着群落参数 c 变化的丰度中位数估计值 μ，对应 ZIB、ZIN$_c$ 和 ZIN 模型（模拟研究 B），访问次数为 $T = 7$、地点数为 $n = 500$。各子图对应 $\mu = 1$、2 与个体探测概率 $r = 0.25$、0.5 的四种组合。为便于查看，省略了取值较高的数据点。

缓解最大似然估计量的偏差问题。

˜在图4中，发现 ZIB估计量 ψ_0 在 $c > 0$时会低估 ψ，并且与 $\mu = 2$相比，在 $\mu = 1$时观察到更大的偏差。当 $c = 1$ 且 $\mu = 1$时，相对偏差达到 -35% 的最大值。发现命题4的近似公式˜与 ψ_0的变化趋势一致，尽管该公式未在图中展示。根据命题 5，发现ZIN˜占域估计量 ψ_1 在 c 范围的两端均能很好地拟合 ψ。在 $c \approx 0$ 附近观察到正偏差，在 $c \approx 1$附近观察到负偏差。已证明，当 $c \approx 0$时，$\frac{\partial}{\partial c}\tilde{\psi}_1$ 的偏导数为正，表明该区域内 ψ 存在高估。然而，在 $c \approx 1$时，发现偏导数在大多数情况下为正，但在某些情况下观察到负值，例如˜在 $\mu = 2$ 且 $r = 0.15$时。在这种情况下，ψ_1 相对于 c 的图表现出类似于ˆ 拱桥的变化，仅在所有 $0 < c < 1$情况下观察到正偏差。发现ZIN$_c$ 估计量 ψ通常是无偏的，除非 $c \approx 1$。当 μ或 r 增加时，发现偏差会减小。

接下来，我们根据网络表7-12中的结果，更详细地总结三个占据率估计量的表现。网络表 7的上半部分（˜ $\mu \times r = 0.25$和 $T = 5$）显示，ZIB占据率估计量 ψ_0 在$c = 0.25$时的相对偏差为 -11%，并且随着 c 增加，这一偏差变得更加明显，在$c = 0.75$时达到最大值 -31%。然而，如网络表 7-12 所示，增加 T或 $\mu \times r$ 可以降低 ψ_0 的偏差，这与命题 4 一致。在网络表 7–9（$\mu = 1$）中，ZIN 占据率

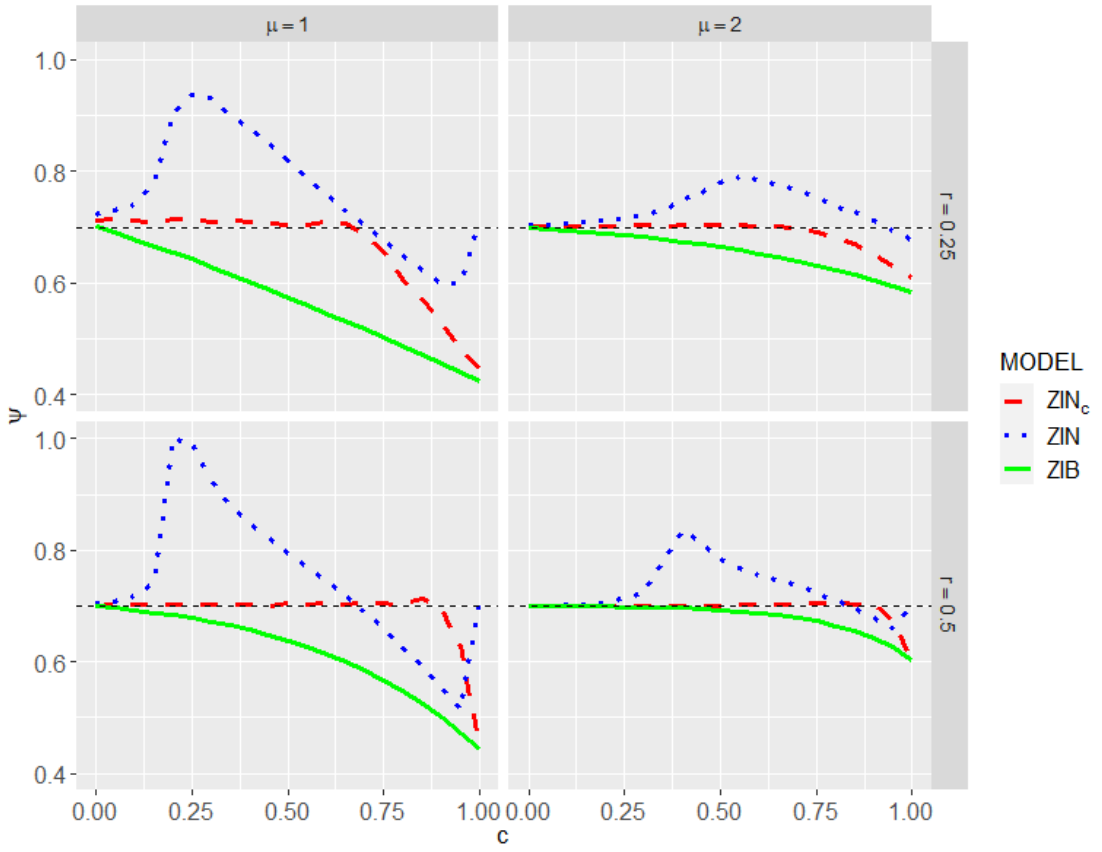


图4：ZIB、ZIN 和ZIN 模型的占据率 ψ 中位数估计值（模拟研究 B）相对于群落参数 c 的绘图，其中访问次数为 $T = 7$ ，地点数量为 $n = 500$。子图对应于 $\mu = 1$、2 以及个体探测概率 $r = 0.25$、0.5 的四种组合。

˜估计量 ψ_1 在 $c = 0.25$的相对偏差约为28%-40%。具体而言，当 $T \geq 7$ 和 $r = 0.5$ 时，ZIN模型经常在 $c = 0.25$ 将 ψ 估计为1（网络表8-9）。~ψ_1 的偏差随着 c增加而降低，在 $c = 0.75$时表现出较小的负偏差。当 μ 增加到~$\mu = 2$时，网络表10-12显示， ψ_1 在所有情况下都高估了 ψ ，在 $c = 0.5$时偏差最为显著，约为10%-13%；而在 $c = 0.25$和0.75时，偏差较小，甚至可以忽略不计。网络ˆ表7-12还显示，ZIN$_c$ 占域估计量 ψ 的偏差通常接近0，但在 $r = 0.25$、 $T = 5$和 $c = 0.75$ 的少数情况下除外（网络表 7）。

ZIB 占据率估计量的 MAD最低，表明它是三种方法中最稳定的。当 $\mu = 1$ 时，ZIN 估计量的 MAD随c 增加而降低；然而，当 $\mu = 2$ 时，由于 ZIN估计量在此处存在显著偏差，$c = 0.5$ 处的 MAD更为突出。相比之下，ZIN$_c$ 估计量的MAD 随 c 值增加而增加；当 $\mu = 1$ 时， $c = 0.75$ 处的MAD 可以是 $c = 0.25$处MAD 的 1.5 到3倍，但当 $\mu = 2$ 时，增幅要小得多。总体而言，所有三个估计量的稳定性都随着 n、 T、 r 和 μ值的增加而提高。

ZIB占据率估计量的中位数标准误与相应的平均绝对偏差较为吻合，但在 $c = 0.75$处观察到了一些负偏差。然而，所得的区间估计量仅在 $\mu = 2$ 和 $c = 0.25$时可靠；在其他情况下，相应的覆盖概率甚至可能达到零，表明其性能非常差。对于ZIN占据率估计量，中位数标准误往往高估平均绝对偏差，尤其是随着 c的增加而更加明显，但当 r 或 μ 增加时，这一问题会有所缓解。所得的覆盖概率也表现出许多异常值，即使在

在某些情况下，它似乎接近名义水平。例如，网络表8-9 显示，覆盖概率在上方各面板中可达到约99%，但在底部各面板中可能降至零。因此，ZIN模型的区间估计不被认为是可靠的。ZIN$_c$ 估计量作为最大似然估计，在数据信息充足时，其在中位数标准误和覆盖概率指标上的表现可能较好，但在其他情况下则无法保证其表现。

最后，我们注意到，丰度估计量同样可能受到违反模型假设或数据不足等问题的影响，在某些情况下可能导致显著偏差，例如当 $c = 0.25$ and $r = 0.25$。一般而言，与丰度估计量相比，相应的占域估计量受这些问题的影响相对较小，性能也相对稳健。

5 示例

5.1 示例 1. Fisher 数据

费舍尔（$Martes\ pennanti$）是原产于北美北方森林的一种食肉哺乳动物。2000年，开展了一项使用非侵入性方法的调查项目，以收集加利福尼亚州北部和中部地区费舍尔物种分布的数据（Zielinskiet al., 2005；Royleand Dorazio, 2008）。这些数据由 $n = 464$ 地点的多次访问记录组成，每个地点访问 $T = 8$ 次。数据见网络表13。在八次访问中，有400个地点的计数为零，因此样本占据率为 $64/464 = 13.8$%。我们使用第2节和第3节中介绍的模型分析了费舍尔数据。结果总结于表1的上部面板中，其中我们还计算了每个模型的赤池信息准则（赤池信息准则）值，以评估模型拟合效果。

ZIN和ZIN$_c$ 模型的赤池信息准则值分别显著低于N–混合模型和N$_c$–混合模型，表明纳入零膨胀概率可以改善模型对费舍尔数据的拟合。在三个零膨胀模型中，ZIN$_c$ 的赤池信息准则值最小。此外，ZIN$_c$ 模型估计群落参数 c 位于(0,1)的中间位置，相关的Wald型置信区间（Web表14）为拒绝原假设 $c = 0$ 和 $c = 1$提供了有力证据。检验 $c = 0$ 和 $c = 1$ 的似然比统计量分别为40.21和8.23，基于10,000个自助样本得到的 p–值分别为0和0.006，从而得出拒绝ZIB和ZIN模型的相同结论。总之，ZIN$_c$ 模型对费舍尔数据的拟合优于其他模型，但三个零膨胀模型的占据率估计值相似。

值得注意的是，ZIN模型估计 ψ 为0.175，这似乎高于其他模型，但ZIN模型的未占据概率为$(1 - \psi) + \psi \exp(-\mu)$。因此，为了与其他模型比较占据率，我们应使用 $\psi\{1 - \exp(-\mu)\}$，其估计值为 $0.175\{1 - \exp(-1.79)\} = 0.146$ （称为ZIN模型的已占据概率），这与其他 ψ的估计值相近。相比之下，所得的丰度估计值差异很大：ZIN给出的值较高，而ZIB的值最低。三个模型的标准误估计值也呈现相同的排序，其中ZIN的参数估计值显示出最大的估计变异。这一现象与模拟研究 B 的网络表8中观察到的情况类似。

模型	μ	r	c	ψ	AIC
			Fisher		
ZIN$_c$	0.93 (0.21)	0.49 (0.10)	0.51（0.13）	0.154（0.020）	626.9
ZIN	1.79 (0.68)	0.22 (0.05)	1*	0.175 (0.030)	633.2
ZIB	0.51 (0.04)	1*	0*	0.140 (0.016)	663.1
N$_c$–混合。	0.13 (0.02)	0.46（0.03）	0.93 (0.02)		636.8
N–混合。	0.16 (0.02)	0.37 (0.02)	1*	0.150 (0.017)	650.6
			冠蓝鸦		
ZIN	48.96（427.97）	0 (0.04)	1*	0.729 (0.091)	170.1
ZIB	0.22（0.03）	1*	0*	0.723 (0.077)	168.1
N–混合。	1.64（0.56）	0.09 (0.03)	1*	0.806 (0.109)	169.5
			灰猫嘲鹪		
ZIN$_c$	0.60 (0.32)	0.42 (0.21)	0.19 (0.14)	0.421（0.079）	139.1
ZIN	0.85（1.31）	0.16（0.09）	1*	0.740 (0.691)	137.8
ZIB	0.26（0.04）	1*	0*	0.403（0.074）	138.5
N$_c$–混合物。	0.52 (0.17)	0.19 (0.06)	0.97 (0.12)		137.9
N–混合。	0.55 (0.14)	0.18 (0.04)	1*	0.421 (0.079)	135.9
			普通黄喉地莺		
ZIN$_c$	0.93 (0.23)	0.47 (0.09)	0.56 (0.10)	0.775 (0.075)	222.9
ZIN	2.06 (0.81)	0.20 (0.05)	1*	0.851 (0.123)	227.8
ZIB	0.49 (0.04)	1*	0*	0.723 (0.064)	254.5
N$_c$–混合。	1.09 (0.38)	0.31 (0.09)	0.88 (0.15)		226.7
N–混合。	1.45 (0.25)	0.23 (0.03)	1*	0.765 (0.059)	226.8
			树燕		
ZIN$_c$	0.75 (0.22)	0.45（0.08）	0.61 (0.12)	0.682 (0.106)	196.4
ZIN	1.81 (0.75)	0.18 (0.05)	1*	0.726 (0.135)	204.0
ZIB	0.40 (0.04)	1*	0*	0.587 (0.071)	228.0
N$_c$–混合。	0.48 (0.11)	0.46 (0.07)	0.72 (0.08)		198.4
N–混合。	1.02（0.19）	0.22 (0.03)	1*	0.641 (0.069)	203.9
			歌带鹀		
ZIN$_c$	1.01 (0.59)	0.37（0.08）	0.90 (0.09)	0.678 (0.240)	184.7
ZIN	2.23 (0.98)	0.21（0.06）	1*	0.573 (0.105)	187.9
ZIB	0.54 (0.05)	1*	0*	0.501（0.072）	210.5
N$_c$–混合。	0.58 (0.13)	0.42 (0.05)	0.93（0.04）		183.4
N–混合。	0.81（0.16）	0.31 (0.04)	1*	0.555 (0.071)	190.5

表1：拟合费舍尔数据和BBS调查中五种鸟类物种数据的各模型的参数估计值（括号内为标准误）及赤池信息准则值。请注意，* 表示该参数值为固定常数，无需估计。N–混合模型的占据率 ψ 是一个推导参数，定义为 $\psi = 1 - \exp(-\mu)$；N$_c$–混合模型没有占据率（第2.2节）。对于冠蓝鸦数据，省略了ZIN$_c$ 和N$_c$–混合模型的结果，因为它们分别简化为ZIB模型和N–混合模型。五种鸟类物种按照ZIN$_c$ 模型中 c 的估计值排序。

5.2 示例2：繁殖鸟类调查（BBS）数据

此示例旨在了解同一调查数据中的群落参数 c 如何因不同物种而异。我们考虑 Royle（2006）使用的数据，这些数据最初来自北美繁殖鸟类调查（BBS）项目，其中包含5种鸟类的出现数据记录——冠蓝鸦（*Cyanocitta cristata*）、猫鹊（*Dumetella carolinensis*）、普通黄喉地莺（*Geothlypis trichas*）、树燕（*Tachycineta bicolor*）和歌带鹀（*Melospiza melodia*）——来自50个地点，每个地点重复访问11次。这5种鸟类的样本占据率（按上述顺序）分别为66%、38%、72%、58%和52%；数据见网络表13。

结果见表1，其中物种的顺序实际上是根据 ZIN$_c$ 模型中 c 的估计值排序的。前两个物种的 c估计值较小；对于冠蓝鸦数据，由于 $c = 0$的估计值，ZIN$_c$ 模型甚至退化为ZIB模型。对于这两个物种，所有考虑的模型都显示出相近的赤池信息准则值，表明模型拟合程度相似。然而，每个模型的 μ、 r和 c 参数估计值产生了相互矛盾的结果。我们认为这是模型可辨识性问题，与Royle（2006）和Link（2003）所描述的问题类似。Royle（2006）得出结论，当探测概率较低时，占据模型中的这一问题更为明显，而这正是这两个物种的情况。

接下来的两个物种——普通黄喉地莺和树麻雀——的 c估计值居中，其95% Wald型置信区间均不包含0或1（网络表14）。此处的模型比较结果可概括为与示例1相似，尤其是零膨胀模型。根据赤池信息准则，ZIN$_c$ 模型的表现优于其他模型。

最后一个物种歌带鸦的 c 估计值较高，为0.90，接近1。事实上，其95%置信区间的上限超过了1的上界。在这种情况下，ZIN$_c$ 模型中$\hat{}\mu$ 和 ψ 的标准误相对较高，表明随着 c 趋近于1，估计的不确定性会增加。模拟研究 B中也观察到了这一现象。

在网络表14中，我们还报告了似然比检验的bootstrap p–值，用于检验ZIB和ZIN模型的假设；结果表明，对于BBS数据中最后三个物种，有显著证据反对零假设。

6 讨论

N–混合模型已扩展为允许样本地点的个体数量在每次访问时发生变化。在这一扩展模型中，每次访问时地点的个体数量被视为一个潜变量，并分解为两个组成部分，而原始 N–混合模型仅考虑其中一个部分。需要注意的是，由于假设 μ恒定，扩展后的 N$_c$–混合模型仅适用于封闭种群（单个季节）。虽然可以通过使用地点和/或访问协变量的回归模型来放宽这一假设，但这种方法是否适用于某些多季节数据尚不明确。

与 N–混合模型不同，N$_c$–混合模型的占据率可能无法通过丰度参数明确定义。为解决这一问题，我们提出了一种零膨胀 N$_c$–混合模型，使用零膨胀概率来明确界定占据率。这一新模型将常用的标准占据率模型和零膨胀 N–混合模型统一到一个框架中。因此，我们发现了这两种模型的一些有趣特性，例如在零膨胀 N–混合模型下使用标准占域估计量，反之亦然。

我们的扩展模型在拟合数据时更加灵活，但也增加了模型的复杂性。这些模型还可能遇到数值不稳定、参数可识别性和模型识别问题，尤其是在数据稀疏时，例如检测概率低、丰度低、访问次数的重复次数少，或地点样本量小的情况。在模拟研究和数据分析中都观察到了这些问题。事实上，最初在 N–mixture 模型中就提出了这些问题（Dennis 等，2015年；Barker 等，2018；Link 等，2018），以此警示人们谨慎使用这些模型。我们的扩展模型需要额外的群落参数，这可能会加剧这些问题。

这些问题。不过，通过纳入额外的捕获-再捕获数据（Barker 等，2018）或在模型中加入协变量（Link 等，2018），可以缓解这些问题。还需要进一步研究，以改进这些情境下的估计。

有几种可能的途径可以扩展这项工作，而每一种途径都需要进一步的实现和研究。

- 使用负二项分布而非泊松分布对丰度进行建模。这种分布通常更适用于生物学与生态学数据分析，但还包含一个额外的聚集参数，从而增加了模型的复杂性和估计难度。

- 考察计数数据而非出现数据。这将需要一个没有显式形式的似然函数，使估计在计算上非常繁重，尤其是对于大型数据集。

- 将协变量纳入模型。纳入协变量对于实际应用非常重要。可以通过对丰度使用对数链接，以及对探测和占域概率使用逻辑斯蒂链接来实现。在联合物种模型中，将物种协变量纳入群落参数也可能具有意义。

- 扩展至检测时间数据。近期已有检测时间占域模型问世，相较于出现数据模型也展现出一些优势（Priyadarshani 等，2022）。Strebel 等（2021）提出了一种 N–混合检测时间占域模型，无需标记个体即可估算丰度。目前正在开发一种用于检测时间数据的类似 N_c–混合占据模型。

致谢

本研　究得到了台湾国家科学及技术委员会的资助　　.

参考文献

Barker, R.J.、Schofield, M.R.、Link, W.A. 和 Sauer, J.R.（2018）。关于用于计数数据的 N–mixture 模型的可靠性。《生物计量学》，74，369–377。Dail, D. 和 Madsen, L.（2011）。用于从开放式超种群的重复计数中估计丰度的模型。《生物计量学》，67，577–587。D´enes, F.V.、Silveira, L.F. 和 Beissinger, S.R.（2015）。估计未标记动物种群的丰度：考虑不完全探测及零膨胀的其他来源。《生态学与进化方法》，6，543–556。Dennis, E.B.、Morgan, B.J. 和 Ridout, M.S.（2015）。N–mixture 模型的计算方面。《生物计量学》，71，237–246。Duarte, A.、Adams, M.J. 和 Peterson, J.T.（2018）。将 N–mixture 模型拟合到具有未建模异质性的计数数据：偏差、诊断和替代方法。《生态建模》，374，51–59。

Gomez,J.P., Robinson, S.K.,Blackburn,J.K.和Ponciano, J.M.（2018年）。用于多物种丰度估计的N–混合模型的高效扩展。《生态学与进化方法》，9，340–353。Haines,L.M.（2016a）。关于重复探测–未探测数据的Royle–Nichols 模型的注记。《农业、生物与环境统计学杂志》，21，588–598。Haines,L.M.（2016b）。N–混合模型的最大似然估计。《生物计量学》，72，1235–1245。Hayes,D.B.和Monfils, M.J.（2015年）。鸟类定点计数的占据建模：活动动物的影响。《野生动物管理杂志》，79，1361–1368。Hogg, S.E., Wang,Y.和 Stone, L.（2021）。存在不完全探测时联合物种分布模型的有效性。《生态学与进化方法》，12，1458–1474。Johnson, D.S., Conn, P.B., Hooten, M.B., Ray, J.C. 和 Pond, B.A.（2013）。大型数据集的空间占据模型。《生态学》，94，801–808。Joseph, L.N., Elkin, C., Martin, T.G,和Possingham, H.P（2009）。使用N–混合模型对丰度进行建模：考虑生态机制的重要性。《生态学应用》，19，631–642。Karavarsamis, N. 和 Huggins, R.（2020）。占据数据分析的两阶段方法 I：齐性情形。《统计学通信——理论与方法》，49，4751–4761。Ke, A., Sollmann,R., Frishkoff,L.O. 和Karp, D.S.（2022年）。用于估计行为变异的层级N–混合模型及新热带鸟类案例研究。《生态学应用》，32，e2632。Kendall,W.L. 和While, G.C.（2009）。关于在地点占据研究中以空间子单元替代重复时间抽样的警示性说明。《应用生态学杂志》，46，1182–1188。Kendall,W.L., Hines, J.E., Nichols, J.D.和Grant,E.H.C.（2013）。放宽占据模型中的封闭性假设：交错的到达和离开时间。《生态学》，94，610–617。Link,W.A.（2003）。具有异质探测概率的捕获-再捕获数据无法识别种群规模。《生物计量学》，59，1123–1130 Link, W.A., Schofield,M.R.,Barker,R.J. 和Sauer, J.R.（2018年）。关于N–混合模型的稳健性。《生态学》，99，1547–1551。MacKenzie, D.I., Nichols, J.D., Lachman, G.B., Droege,S., Royle, J.A.和Langtimm,C.A.（2002）。当探测概率小于1时估计地点占据率。《生态学》，83，2248–2255。MacKenzie, D.I., Nichols, J.D., Seamans, M.E. 和 Guti´errez, R.J.（2009）。使用多个状态和不完全探测对物种出现动态进行建模。《生态学》，90，823–835。MacKenzie, D.I., Nichols, J. D., Royle, J.A., Pollock, K.H.,Bailey,L.L.和Hines, J. E.（2017）。占据率估计与建模：推断物种出现的模式和动态，第2版。伦敦：Elsevier/Academic Press。

Otto, C.R.V., Bailey, L.L., and Roloff, G.J.（2013）。当采样违反封闭性假设时，改进物种占据率估计。《Ecography》，36，1299–1309。Priyadarshani, D., Altwegg, R., Lee, A.T.K., and Hwang, W.H.（2022年）。占据模型能从检测时间数据中获得什么？《Ecology》，103，e3832。R Core Team（2022年）。R：一种用于统计计算的语言和环境。R Foundation for Statistical Computing，奥地利维也纳。Royle, J.A. and Nichols, J.D.（2003）。利用重复出现-缺失数据或样点计数估计丰度。《Ecology》，84，777–790。Royle, J.A.（2004）。用于根据空间重复计数估计种群大小的N–混合模型。《生物计量学》，60，108–115。Royle, J.A.（2006）。探测概率存在异质性的地点占据模型。《生物计量学》，62，97–102。Royle, J.A. and Dorazio, R.M.（2008）。生态学中的层次建模与推断：种群、集合种群和群落数据分析。Elsevier/Academic Press。Self, S.G., and Liang, K.L.（1987）。非标准条件下最大似然估计量和似然比检验的渐近性质。《Journal of the American Statistical Association》，82，605–610。Strebel, N., Fiss, C.J., Kellner, K.F., Larkin, J.L., K´ery, M., and Cohen, J.（2021）。基于检测时间数据估计丰度。《生态学与进化方法》，12，909–920。Zielinski, W.J., Truex, R.L., Schlexer, F.V., Campbell, L.A., and Carroll, C.（2005）。美国加利福尼亚州内华达山脉森林中食肉动物的历史分布与当代分布。《Journal of Biogeography》，32，1385–1407。