

50-Year-Old Curiosities: Ancillarity and Inference in Capture–Recapture Models

Matthew Schofield and Richard Barker

Abstract. We review developments from the late 1950s, starting with the work of John Darroch, that led to the models of Cormack [*Biometrika* **51** (1964) 429–438], Jolly [*Biometrika* **52** (1965) 225–247] and Seber [*Biometrika* **52** (1965) 249–259] that are commemorated in this volume. We emphasize some of the fundamental contributions that were pivotal and often ahead of their time. We look at how these early contributions helped to shape the field and illustrate important concepts in statistics, including sufficiency, ancillarity, partial likelihoods, missing data and model fitting in the presence of latent variables. We also identify two curiosities. The first is the long-held and mistaken belief that the maximum likelihood estimators for various capture–recapture models are in common. Using various notions of ancillarity, we show that the maximum likelihood estimators from partial models (like the Cormack–Jolly–Seber model) will in general differ from full likelihood approaches (such as the Jolly–Seber model). The second is the belief that model specification in terms of latent variables is a relatively recent advance. We highlight how Jolly (1965) used a state-space model to describe the problem, using latent variables to separate the capture and mortality processes. We show how Markov chain Monte Carlo can be used to fit this model and how it relates to other capture–recapture models specified in terms of latent variables.

Key words and phrases: Ancillarity, capture–recapture, data augmentation, goodness of fit, sufficiency.

1. INTRODUCTION

The basis for modern capture–recapture analysis was provided by the models of Cormack (1964), Jolly (1965) and Seber (1965) that are commemorated in this volume. These models had their genesis in closed-population theory which has a long history. The foundation for modern open-population models was established by Darroch (1959), who in turn built on earlier contributions by Fisher and Ford (1947) and Leslie, Chitty and Chitty (1953), among others. Joan Fisher Box in her biography of her father provides this enig-

matic quote: “In the early month of 1939, during Jackson’s third visit, it occurred to Fisher that a backward and forward progression (of capture–recapture data) would supply separately the rates of birth or immigration and of death or emigration” (Box, 1978, page 303). Fisher’s idea was implemented in Jackson (1939) and it thus appears that Fisher anticipated the dual forward and backward in time formulation that was fully described nearly 60 years later by Pradel (1996). See also Nichols (2016) in this volume.

We begin by overviewing the development of early open-population models, starting with Darroch (1959) and ending with the contributions by Jolly (1965) and Seber (1965). Our focus is on the underlying probability models, how they are factored, what unobserved random variables are included as parameters, and what quantities are conditioned on. It is these features that distinguish among the various models. Key

Matthew Schofield is a Lecturer, Richard Barker is Professor of Statistics, Department of Mathematics and Statistics, University of Otago, PO Box 56, Dunedin 9054, New Zealand (e-mail: mschofield@maths.otago.ac.nz; rbarker@maths.otago.ac.nz). The authors contributed equally to this manuscript.

TABLE 1
Notation

Term	Definition
<i>Fixed constants</i>	
k	Number of capture occasions.
<i>Parameters and latent variables</i>	
N_i	Population size at the time of sample i ($i = 1, \dots, k$).
ν	The number of individuals that were ever available for capture in at least one sample.
U_i	Number of unmarked individuals in the population at the time of sample i .
M_i	Number of marked individuals in the population at the time of sample i .
M_{ij}	Number of individuals released in sample i still alive and available for capture in sample j ($i = 1, \dots, k-1, j = i+1, \dots, k$).
p_i	Probability that an individual is caught in sample i .
ϕ_i	Probability that an individual in the population at the time of sample i is still in the population at the time of sample $i+1$.
B_i	The number of individuals entering the population between the time of sample i and the time of sample $i+1$. B_0 refers to the number at the start of the study.
β_i	Probability that an individual enters the population between samples i and $i+1$ ($i = 0, \dots, k-1; \sum_{i=0}^{k-1} \beta_i = 1$) where β_0 is the probability an individual is present at the start of the study.
η_i	Probability that an individual caught in sample i is removed (i.e., become a “loss on capture”).
<i>Statistics</i>	
x_ω	Number of individuals with capture history $\omega \in \Omega$. The value x_{11001} reflects the number of individuals caught in the first, second and fifth sampling occasion of a $k = 5$ study. The number $x_{00\dots 0}$ is not observable.
\mathbf{x}^{obs}	Collection of the numbers of individuals with each observable capture history.
m_{ij}	Number of individuals released in sample i recaptured in sample j .
u_i	Number of unmarked individuals caught in sample i .
n	Number of individuals caught at least once in the study.
m_i	Number of marked individuals caught in sample i .
R_i	Number of individuals released in sample i .
r_i	Number of individuals released in sample i that were ever recaptured again.
z_i	Number of marked individuals not caught in sample i but recaptured after i .

to understanding these differences are the fundamental concepts of sufficiency, ancillarity, partial likelihood and latent variables. This reveals curious differences, including that the maximum likelihood estimators (MLEs) for the Cormack–Jolly–Seber (CJS) model do not correspond to MLEs for the Jolly–Seber (JS) model, despite widely held beliefs to the contrary.

Throughout the manuscript we use notation as defined in Table 1. Occasionally we follow a citation that uses nonstandard notation: in such cases we define the notation inline. We use bold letters such as \mathbf{y} to represent collections such as y_1, \dots, y_n .

2. DARROCH, CORMACK, JOLLY AND SEBER

2.1 Darroch 1959

The model of Darroch (1959) was an extension of his closed-population model (Darroch, 1958) to allow for animals entering and leaving the population during the study. The closed population models \mathbf{x}^{obs} as multinomial observations with index ν and probabilities π_ω

derived as functions of the capture probabilities p_i . In the notation of Table 1 this model can be written as

$$(2.1) \quad f(\mathbf{x}^{\text{obs}} | \nu, \mathbf{p}) = \frac{\nu!}{\prod_\omega x_\omega!} \prod_\omega \pi_\omega^{x_\omega}.$$

Darroch referred to this as model A.

The difficulty in extending model A to open populations lay in the fact that on each sampling occasion the number of individuals in the population differs according to the combined effect of additions and losses to the population.

Allowing immigration only, Darroch extended (2.1) by setting the abundance at the time of the first sample N_1 equal to ν and introduced the births \mathbf{B} as parameters. He was then able to find simple expressions for the MLEs for the model.

Allowing death only, Darroch introduced the parameters ϕ leading to a straightforward extension of the model (2.1) in which the π_ω are now functions of the p_i and ϕ_i . For this model he was also able to find explicit formulae for the MLEs.

For the general case of immigration and death, Darroch (1959) was unable to find simple expressions for the MLEs. The stumbling block for Darroch was parameterizing the model in terms of N_i . The consequence was an intractable likelihood, expressed in terms of a generating function comprising $2^{k(k+1)/2}$ terms, each term itself a product of probabilities. For a $k = 7$ study there are ca. 268 million of these terms. Darroch did point the way to an alternative method based on equating some class sizes to their expected values and he was able to solve this for the $k = 4$ case. This technique featured in subsequent work by Darroch's Ph.D. student George Seber.

2.2 Seber 1962

Seber (1962) studied the case where a number a_i of batch-marked animals are added to the population in sample i . A sample of size b_i is then taken, after allowing time for mixing. This sample is permanently removed from the population (e.g., it might represent a commercial harvest). The b_i contain both marked and unmarked animals and which batch an individual came from can be determined by the distinguishing marks. Also, the model allows for marked animals to die in the interval between these samples, both indexed by i .

This model has two notable features. The first is that the releasing of the a_i marked animals occurs separately from the subsequent removal sampling of b_i individuals. Seber allowed mortality to occur between samples a_{i-1} and b_{i-1} , as well as between sample b_{i-1} and the release of a_i . The parameters describing these two mortality events cannot be separately estimated. The confounded survival parameter (the probability of surviving from a_{i-1} to a_i) can be expressed in the now standard notation as ϕ_{i-1} .

The second feature was that Seber included a binomial term for the captures of unmarked animals, under the assumption that all individuals in the population at the time of sample b_i were subject to the same capture probability p_i . The index of each binomial term was U_i , and Seber's innovation was to regard these as unknown parameters. If we use \mathbf{x}^{obs} to represent the raw batch mark data with summaries \mathbf{m} representing recaptures of marked animals, and \mathbf{u} to denote the number of captured unmarked animals (from a total \mathbf{U}), then Seber's model can be written as

$$\begin{aligned} f(\mathbf{x}^{\text{obs}}|\mathbf{U}, \mathbf{p}, \boldsymbol{\phi}) &= f(\mathbf{x}^{\text{obs}}, \mathbf{u}, \mathbf{m}|\mathbf{U}, \mathbf{p}, \boldsymbol{\phi}) \\ &= f(\mathbf{x}^{\text{obs}}|\mathbf{m}) f(\mathbf{m}|\mathbf{u}, \mathbf{p}, \boldsymbol{\phi}) f(\mathbf{u}|\mathbf{U}, \mathbf{p}). \end{aligned}$$

In providing his likelihood, Seber (1962) did not include the term $f(\mathbf{x}^{\text{obs}}|\mathbf{m})$ which does not contain any

parameters, although this can be used for assessment of goodness of fit, a point we return to later.

The algebraic structure of Seber's (1962) model for recaptures conditional on the releases is represented above by the term $f(\mathbf{m}|\cdot)$. This structure coincides with what became known as model M_1 of Brownie et al. (1976), although this is not generally recognized except by Seber himself (Seber, 1970, page 314; see also the comment by Jolly 1965, discussed below). We also now know that model M_1 is algebraically identical to the model of Cormack (1964).

The full generality of Seber's (1962) model was not obvious at the time because it was developed in terms of dead recovery or band recovery data, that is, data on individuals that can only be recaptured once. As discussed by Brownie et al. (1985) (Section 8.2) and Burnham (1991), the theories for band recovery and live-recapture data developed separately even though the basic model describing the data is the same. This was not fully understood in the early 1960s, although Jolly pointed out that the estimators from Seber (1962) were identical to those for the full JS model and that "there were no essential differences between the multiple and the single recapture method of sampling a population" (Jolly, 1965, page 226). Seber himself did not recognize the connections between the two models until he wrote his 1970 book (see the Seber interview in the present volume).

2.3 Jolly 1963

Adopting what is now standard notation, Jolly (1963) provided maximum likelihood solutions to the open-population problem with both immigration and death, but assumed a deterministic model for births and deaths in which parameters such as survival probability and immigration rate are replaced by ratios of adjusted population size (adjustments including allowance for loss on capture). Jolly derived estimators of these quantities, including the abundances, and showed that these could be expressed in terms of estimates of the latent variables M_i , and the usual summary statistics. These estimators are identical to those given subsequently in Jolly (1965), the difference being how they were motivated.

The likelihood used by Jolly (1963) was based on multinomial distributions that condition on $m_i + u_i$, the number caught in each sample, commonly denoted n_i . Jolly realized that this was not strictly correct, as the sample size is not fixed in practice, but reasoned that variability in $m_i + u_i$ would have little effect on inference under his model. As we will discuss in the next

section, this can be considered an appeal to near ancillarity, a concept that can justify conditioning used in subsequent developments, including the model of Cormack (1964).

2.4 Cormack 1964

Cormack (1964) observed that all methods developed to date focused on estimating population size, and assumed that animals captured on each sampling occasion were a random sample of the whole population. Motivated by a mark-resighting study of fulmar petrels, Cormack (1964) introduced his model in which the focus of inference was on survival probabilities. In the fulmar study there were two separate sampling processes, one involving the capture of unmarked birds and the other involving resightings of marked individuals. Since the sole function of the capture of unmarked birds was to provide individuals for study in future years, Cormack reasoned that the number of such releases could be taken as fixed. By this he meant that they could be treated as though they were predetermined by the researcher even if they were not.

Cormack (1964) extended model A of Darroch (1958) to open populations, conditioning on the numbers released on each occasion. He found the MLEs for the ϕ_i and p_i for this model, now the standard notation for the survival probability and capture probability respectively. These equations were in fact identical to those found for the deterministic survival and sampling fractions denoted μ and f by Jolly (1963). Whether Cormack was aware of this is unclear, as no reference is made to this in Cormack (1964). However, Cormack had arrived at his estimates through a fully specified random sampling model up to conditioning on the numbers released (which is equivalent to conditioning on the \mathbf{u}). Thus, Seber (1962) and Cormack (1964) had now identified a key part of the full probability model: that of the model for recaptures conditioned on the numbers released, the term $f(\mathbf{m}|\mathbf{u}, \mathbf{p}, \phi)$ in Seber's (1962) model.

2.5 Jolly (1965) and Seber (1965)

Jolly (1965) described a general model that conditions on the number in the population at the time of sample 1, and parameterized in terms of numbers of births B_i , capture probabilities p_i , survival probabilities ϕ_i and parameters η_i regulating losses on capture. Starting immediately before the first sample, Jolly models the progression of an individual through the study. At the time of sample i , individuals are divided into survivors from previous occasions or newly joined

individuals. These are then subject to a capture process at time i , and then a survival process to determine if they are present at the time of sample $i + 1$. The model includes the random variables B_i ($i = 0, \dots, k - 1$), U_i ($i = 1, \dots, k$), and M_{ij} ($i = 1, \dots, k - 1$, $j = i + 1, \dots, k$), the unknown number of individuals last caught in occasion i and still alive in j . Thus, Jolly (1965) describes a joint model for the capture history data as well as for unmarked individuals that is factorized into a term for captures of unmarked animals conditional on the number of unmarked individuals in the population, a term for survival of individuals, whether marked or unmarked, and a term for losses on capture.

To derive the likelihood, Jolly noted that we must either marginalize over the unobserved variables or treat them as parameters. Attempting to marginalize leads to the problems encountered by Darroch (1959). Treating latent variables as parameters was the key to solving the problem and led Jolly to the now familiar estimators for \hat{N}_i , \hat{B}_i , $\hat{\phi}_i$, \hat{p}_i and $\hat{\eta}_i$ and their covariances.

Seber (1965) took a slightly different route but obtained the same estimators. One difference from Jolly (1965) was that Seber marginalized across some variables (the M_{ij} terms of Jolly) while treating the others as parameters. The result was a different representation of Seber's (1962) model, but now expressed in terms of live recaptures and without the component for losses on capture included by Jolly (1965).

What is now commonly referred to as the Jolly–Seber (JS) model follows Seber (1965) in marginalizing across M_{ij} , leading to

$$(2.2) \quad \begin{aligned} f(\mathbf{x}^{\text{obs}}|\mathbf{U}, \phi, \mathbf{p}, \eta) \\ = f_1(\mathbf{u}|\mathbf{U}, \mathbf{p}) f_2(\mathbf{R}|\mathbf{u}, \eta) f_3(\mathbf{x}^{\text{obs}}|\mathbf{R}, \phi, \mathbf{p}), \end{aligned}$$

the term f_1 representing the model for captures of unmarked individuals, f_2 the model for losses on capture, and f_3 representing what is now known as the Cormack–Jolly–Seber (CJS) model. The term f_3 is equivalent to $f(\mathbf{m}|\mathbf{u}, \mathbf{p}, \phi)$ described previously for Seber's (1962) model when there is no loss on capture.

Jolly's (1965) paper is a *tour de force*. In addition to solving the general problem and providing intuitive explanations for the estimators, he

- Explains the relationship between his model and the earlier model of Seber (1962) as well as Seber (1965). He also showed the relationship between his 1965 estimates and those of Jolly (1963).
- Specified the model in terms of latent variables that separately describe the demographic and sampling processes. Indeed, his Table 1 is an early example

of a graphical description of a state-space model. The inherently Markovian nature of the underlying demographic model for marked individuals was also recognized by Seber (1965, page 251). Latent variable representations of capture–recapture models have dominated more recent literature with the advent of modern statistical computation as we describe later.

- Suggested a way of improving a capture–recapture study by supplementing the recapture statistics with resightings of marked animals taken by “untrained persons over a wide area.” The full model for this generalization was described by Barker (1997) who also showed that Jolly’s intuition was correct. At the same time Jolly anticipated the joint live recapture and dead recovery model of Burnham (1993).

2.6 An Incomplete Selection of Subsequent Developments

Following Jolly (1965) and Seber (1965), there was extensive development of capture–recapture models (see Seber, 1982, 1986, 1992; Schwarz and Seber, 1999, for reviews). Much of the focus of this work was on generalizing the term f_3 in (2.2), for example, to allow for covariates of recapture and survival (Lebreton et al., 1992; Schwarz, Schweigert and Arnason, 1993; Brownie et al., 1993; Bonner and Schwarz, 2006; King et al., 2006) and joint models that include dead-recoveries (Burnham, 1993) or dead recoveries and resightings (Barker, 1997). Other key developments include the robust design of Kendall, Pollock and Brownie (1995), Pollock (1982) and the population growth rate model of Pradel (1996).

Kendall, Pollock and Brownie (1995) conditioned on \mathbf{u} in developing their likelihood for the robust design, noting a slight loss in efficiency. Pradel (1996) conditioned on n , the total number caught during the experiment. The idea of conditioning on aspects of the data has been present throughout the development of capture–recapture theory, but needs to be better understood given that conditioning potentially leads to loss of information.

Especially important for subsequent development was the representation of the model given by Schwarz and Arnason (1996), building on Crosbie (1979) and Crosbie and Manly (1985). Schwarz and Arnason pointed out that the JS model was unsatisfactory, as the births are not explicitly modeled. They express the model by conditioning on v , the number of individuals ever available for capture instead of \mathbf{U} . Given v , the births can be modeled as multinomial with index v and

probability β . Subsequently, the model considered differs in how the unmarked individuals are caught, but the factors f_2 and f_3 from (2.2) are retained:

$$(2.3) \quad \begin{aligned} f(\mathbf{x}^{\text{obs}}|v, \beta, \mathbf{p}, \phi, \eta) \\ = f_0(n|v, \beta, \mathbf{p}, \phi) f_1(\mathbf{u}|n, \beta, \mathbf{p}, \phi) \\ \times f_2(\mathbf{R}|\mathbf{u}, \eta) f_3(\mathbf{x}^{\text{obs}}|\mathbf{R}, \mathbf{p}, \phi). \end{aligned}$$

The factorization (2.3) is appealing, as it fully generalizes the Darroch (1958) model A to open populations, makes clear the conditioning involved in each factor, is expressed in terms of parameters that admit simple and reasonable constraints, includes the CJS model as a special case, and is readily generalizable, for example, to incorporate auxiliary data, or to multiple states, or to the robust design. We subsequently refer to (2.3) as the Crosbie–Manly–Schwarz–Arnason (CMSA) model.

3. CONDITIONING, ANCILLARITY AND SUFFICIENCY

All models that we have considered condition on random variables. The JS and CMSA approaches treat unobserved random variables \mathbf{U} and v respectively as parameters that are estimated. The approaches of Cormack (1964), Kendall, Pollock and Brownie (1995) and Pradel (1996) condition either on observed random variables \mathbf{u} or n . To understand the implications of such conditioning, we consider notions of partial ancillarity as recently done for continuous time closed-population capture–recapture models by Barker et al. (2014).

3.1 Definitions of Ancillarity

An ancillary statistic is one that does not depend on unknown parameters. Ancillary statistics were first considered by Fisher (1925, 1934) in order to improve inference of the model parameters in some situations. Our goal differs from that of Fisher. We wish to use ancillarity to justify the removal of conditional model components that contain nuisance parameters. To do that, we use two definitions of ancillarity: S-ancillarity and Bayes-ancillarity.

Consider an example where we observe data x and y with parameters $\theta = (\theta_1, \theta_2)$. Of interest is θ_1 with θ_2 considered a nuisance parameter. As defined by Barndorff-Nielsen (1976), a statistic x is S-ancillary for θ_1 in the presence of θ_2 if we can express the joint distribution

$$f(x, y|\theta) = f(y|x, \theta_1) f(x|\theta_2).$$

It is straightforward to see that the MLE for θ_1 will be identical if we consider only $f(y|x, \theta_1)$ or the full

model $f(x, y|\theta)$. As described by Ghosh, Reid and Fraser (2010), if the joint distribution is expressed as

$$(3.1) \quad f(x, y|\theta) = f(y|x, \theta_1) f(x|\theta),$$

then x can be S-ancillary if and only if there exists a reparameterization of (θ_1, θ_2) to (θ_1, α) such that

$$f(x, y|\theta) = f(y|x, \theta_1) f(x|\alpha).$$

Bayes-ancillarity is a weaker form of ancillarity defined by Severini (1995) that necessitates a prior distribution for θ_2 denoted $f(\theta_2|\theta_1)$. Suppose the joint distribution is expressed as in (3.1). The statistic x is said to be Bayes-ancillary for θ_1 with respect to $f(\theta_2|\theta_1)$ if

$$\int f(x|\theta) f(\theta_2|\theta_1) d\theta_2$$

does not depend on θ_1 . If we also assume a prior distribution for θ_1 , then Bayes-ancillarity of x implies that the posterior distribution for θ_1 found using only $f(y|x, \theta_1)$ is equivalent to the posterior distribution using the full model $f(x, y|\theta)$. A statistic that is S-ancillary is also Bayes-ancillary.

Although Bayes ancillarity presupposes a prior distribution for θ_2 , it has relevance for frequentist inference about θ_1 in the context of use of integrated likelihood as a device for eliminating nuisance parameters (Berger, Liseo and Wolpert, 1999).

3.2 Elementary Example: A Pair of Binomials

To illustrate the concepts of ancillarity, we consider an example that represents the simplest form of capture–recapture. Suppose we carry out an experiment as follows. First we sample the population and mark the x individuals so caught. We then carry out a second experiment and determine the proportion of marked individuals in the population from a sample of size z . As a model, we might consider x as a binomial observation with unknown index v and parameter p , and y as a binomial observation with index z and parameter p . The joint model for x and y is

$$\begin{aligned} f(x, y|v, p) \\ (3.2) \quad &= f_1(x|v, p) f_2(y|z, p) \\ &= \binom{v}{x} p^x (1-p)^{v-x} \binom{z}{y} p^y (1-p)^{z-y}. \end{aligned}$$

The observation x provides one degree of freedom at the cost of one parameter v . Therefore, it might be supposed that all the information about p is provided by f_2 . But this is obviously not the case; the sufficient

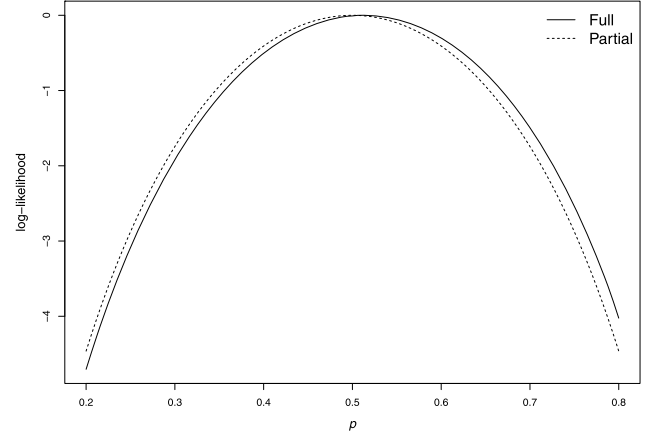


FIG. 1. Profile log-likelihood function for p , obtained from the full likelihood [equation (3.2)], compared to the partial likelihood [which is proportional to f_2 in equation (3.2)] for $x = 20$, $y = 10$ and $z = 20$.

statistic for p is $x + y$, not y since we can rewrite (3.2) as

$$\begin{aligned} f(x, y|v, p) &= \frac{\binom{v}{x} \binom{z}{y}}{\binom{v+z}{x+y}} \\ &\quad \times \binom{v+z}{x+y} p^{x+y} (1-p)^{v+z-x-y}. \end{aligned}$$

Nor is the statistic x S-ancillary for p ; for x to be S-ancillary we must be able to find a reparameterization of (p, v) into (p, α) such that the term f_1 depends only on α and not on p , that is, $f_1(x|\alpha)$. No such reparameterization exists. If we fitted this model by ML, our estimate for p is not the intuitive $\tilde{p} = y/z$, and our estimate for v is not $x/\tilde{p} = xz/y$. Rather, the MLEs are

$$\begin{aligned} \hat{p} &= (x + y)/(\hat{v} + z) \quad \text{and} \\ \hat{v} &= \arg \max_v \frac{v!}{(v-x)!} (1 - \hat{p})^{v+z-x-y}. \end{aligned}$$

To see the difference, consider an example where $x = 20$, $y = 10$ and $z = 20$. The intuitive estimates are $\tilde{p} = 0.5$ and $\tilde{v} = 40$. The MLEs are $\hat{p} = 0.5172$ and $\hat{v} = 38$.

For inference about p we could choose to ignore f_1 in (3.2). In this case, we would be using f_2 as a partial likelihood sensu Cox (1975). Although f_1 does contain information about p , we could justify use of f_2 due to asymptotic equivalence between the full and partial likelihoods (Sanathanan, 1972). In this example, the difference between the partial and full likelihood functions is small (Figure 1). This may not always be the case.

Another justification for ignoring f_1 is Bayes-ancillarity. If we assign v a scale prior,

$$f(v) \propto v^{-1},$$

then

$$\begin{aligned}
 & \sum_v f_1(x|v, p) f(v) \\
 &= \sum_{v=x}^{\infty} \binom{v}{x} p^x (1-p)^{v-x} v^{-1} \\
 (3.3) \quad &= x^{-1} \sum_{U=0}^{\infty} \binom{U+x-1}{U} p^x (1-p)^U \\
 &= x^{-1},
 \end{aligned}$$

which does not depend on p since the summand in line 2 is a negative-binomial probability mass function for the observation $U = v - x$. For inference about p we can use the marginal likelihood after integrating over v . Using (3.3), this is given by

$$\begin{aligned}
 f(x, y|p) &= \sum_v f_1(x|v, p) f_2(y|z, p) f(v) \\
 &= x^{-1} f_2(y|z, p).
 \end{aligned}$$

For inference regarding p , we are justified in using f_2 alone so long as we are willing to use a marginal likelihood for p , found by integrating over v assuming a scale prior. Even if we do not use the scale prior and choose instead some other vague prior, we expect little difference between estimates obtained under the full and integrated likelihoods owing to the close agreement evident in Figure 1.

Now suppose that in our experiment, x is modeled as a Poisson observation with parameter νp . The parameter ν now represents some expectation of the size of the population that was sampled and marked. We can express f_1 as

$$\begin{aligned}
 f_1(x|v, p) &= \frac{e^{-\nu p} (\nu p)^x}{x!} = \frac{e^{-\alpha} \alpha^x}{x!} \\
 (3.4) \quad &= f_1(x|\alpha),
 \end{aligned}$$

where $\alpha = \nu p$. Thus, under the Poisson model the statistic x is S-ancillary for p in the presence of α . The MLE is $\hat{p} = y/z$ and x now conveys no information about p . We can safely ignore the model for x knowing that we can find the MLE for p solely from the information in y . In our example with $x = 20$, $y = 10$ and $z = 20$, the MLEs for the Poisson model correspond to the intuitive estimates: $\hat{p} = 0.5$ and $\hat{\nu} = 40$.

3.3 Ancillarity and Capture–Recapture

3.3.1 Closed-population models. Sandland and Cormack (1984) claimed that the MLE for ν in a multinomial model [equivalent to assuming x has a binomial

distribution in (3.2) above] is equivalent to the MLE for ν in a Poisson model [equivalent to assuming x has a Poisson distribution in (3.4) above]. As seen by the pair of binomials example, this is not true. Later papers by Cormack clarify the relationship. Cormack (1989) noted that two approaches are possible for estimating parameters from a multinomial model: conditional (using only f_2 to estimate the capture probabilities) and unconditional (finding the MLE for the full likelihood). He was uncertain how the MLE from the Poisson model related to these different approaches. Cormack (1992) established that the full Poisson model is equivalent to the conditional multinomial model as we have shown above.

3.3.2 The Jolly–Seber model. The pair of binomials example also serves us well for extension to open-population examples. We start with a slight generalization of the JS model

$$\begin{aligned}
 & f(\mathbf{x}^{\text{obs}}|\mathbf{U}, \boldsymbol{\phi}, \mathbf{q}, \mathbf{p}, \boldsymbol{\eta}) \\
 &= f_1(\mathbf{u}|\mathbf{U}, \mathbf{q}) f_2(\mathbf{R}|\mathbf{u}, \boldsymbol{\eta}) f_3(\mathbf{x}^{\text{obs}}|\mathbf{R}, \boldsymbol{\phi}, \mathbf{p}).
 \end{aligned}$$

Here, we have allowed the capture probability for unmarked animals \mathbf{q} to differ from the capture probability for marked animals \mathbf{p} , which makes sense for the fulmar petrel study of Cormack (1964) since the marking and recapture processes are distinct. We subsequently ignore losses on capture given by f_2 , as this term contains no parameters of interest; that is, the statistics \mathbf{R} are S-ancillary for \mathbf{p} and $\boldsymbol{\phi}$ and need not be modeled once we have conditioned on \mathbf{u} .

The model f_1 assumes that the number of unmarked individuals caught can be modeled as independent binomials, each with index U_i ,

$$f_1(\mathbf{u}|\mathbf{U}, \mathbf{q}) = \prod_{i=1}^k \binom{U_i}{u_i} q_i^{u_i} (1 - q_i)^{U_i - u_i}.$$

Since $q_i \neq p_i \forall i$, the statistics \mathbf{u} are S-ancillary for \mathbf{p} and $\boldsymbol{\phi}$. Thus, Cormack (1964) was justified in conditioning on \mathbf{u} and the MLEs he would have obtained are the same as he would have got from the full model. However, if the marking and recapture processes depend on the same capture parameters (i.e., $\mathbf{q} = \mathbf{p}$), then the statistics \mathbf{u} are not S-ancillary and the MLE for \mathbf{p} and, subsequently, $\boldsymbol{\phi}$ will differ between the JS and CJS models. We do not believe this is commonly understood, as the estimators given by Seber (1965) and Jolly (1965) are usually referred to as MLEs for both the JS as well as the CJS models (see, for example, Pollock et al., 1990, Williams, Nichols and Conroy,

2002). The differences may not be trivial, as is easily demonstrated numerically (see supplementary materials, Schofield and Barker, 2016, Section 2).

The statistic \mathbf{u} is not S-ancillary, but it is Bayes-ancillary for \mathbf{p} and $\boldsymbol{\phi}$ with respect to independent scale prior distributions for U_i :

$$f(U_i) \propto \frac{1}{U_i}, \quad i = 1, \dots, k.$$

The proof follows that used in the pair of binomials example. A consequence is that we are justified in using the CJS model in place of the JS model if we are willing to use a marginal likelihood for \mathbf{p} and $\boldsymbol{\phi}$ having integrated out \mathbf{U} using these independent scale priors. We note that assuming prior independence of \mathbf{U} may not be biologically reasonable, as it is likely we would expect some prior dependence between U_i and U_j , $i \neq j$.

Both Seber (1962, 1965) and Jolly (1965) treat the random variables \mathbf{U} as parameters in key steps of the model derivation. To assess the implications of this for the JS model, we assume that U_1, \dots, U_k have independent Poisson distributions with rate parameters μ_1, \dots, μ_k . Marginalizing over \mathbf{U} leads to an independent Poisson model for \mathbf{u} with rates $p_i \mu_i$,

$$f_1(\mathbf{u}|\boldsymbol{\mu}, \mathbf{p}) = \prod_{i=1}^k \frac{e^{-p_i \mu_i} (p_i \mu_i)^{u_i}}{u_i!},$$

in place of the independent binomial model for \mathbf{u} in the JS model. Adopting the Poisson model leads to S-ancillarity of u_i in the presence of μ_i . This is an interesting result. It means that modeling \mathbf{U} as independent Poisson random variables leads to identical MLEs for \mathbf{p} and $\boldsymbol{\phi}$ to those in the CJS model, but different MLEs to those from the JS model (supplementary materials, Schofield and Barker, 2016, Section 2).

3.3.3 The CMSA model. Similar results to those described above hold for the CMSA model (2.3). Schwarz and Arnason (1996) derive the model for unmarked individuals by first specifying a multinomial distribution for all unmarked individuals ever available for capture that we label as \tilde{f}_1 ,

$$(3.5) \quad \tilde{f}_1(\mathbf{u}|\nu, \boldsymbol{\beta}, \boldsymbol{\phi}, \mathbf{p}) = \frac{\nu!}{u_1! u_2! \cdots u_k! (\nu - n)!} \left(1 - \sum_{j=1}^k \psi_j p_j\right)^{\nu - n} \times \prod_{i=1}^k (\psi_i p_i)^{u_i},$$

where $n = \sum_{j=1}^k u_j$ is the observed sample size and $\psi_1 = \beta_0$, with $\psi_{i+1} = \psi_i(1 - p_i)\phi_i + \beta_i$, $i = 1, \dots, k - 1$. The components f_0 and f_1 in (2.3) are factorizations of this multinomial distribution:

$$(3.6) \quad f_0(n|\nu, \mathbf{p}, \boldsymbol{\beta}, \boldsymbol{\phi}) = \binom{\nu}{n} \pi^n (1 - \pi)^{\nu - n},$$

$$(3.7) \quad f_1(\mathbf{u}|n, \boldsymbol{\beta}, \boldsymbol{\phi}, \mathbf{p}) = \frac{n!}{u_1! u_2! \cdots u_k!} \prod_{i=1}^k \xi_i^{u_i},$$

where $\pi = \sum_{j=1}^k \psi_j p_j$ and $\xi_i = \psi_i p_i / \sum_{j=1}^k \psi_j p_j$, $i = 1, \dots, k$.

Care is needed when establishing ancillarity results for the CMSA model because the parameter vector $\boldsymbol{\beta}$ has only $k - 1$ free parameters since $\sum_{i=0}^{k-1} \beta_i = 1.0$. Any reparameterization we consider in an effort to determine S-ancillarity must satisfy this constraint.

We first compare the model where we condition on \mathbf{u} (i.e., the CJS model) to the model where we condition on n as in Pradel (1996) and Link and Barker (2005). This involves establishing whether any information about \mathbf{p} and $\boldsymbol{\phi}$ is lost by ignoring the model component f_1 . We follow Link and Barker (2005) and consider a reparameterization of $\boldsymbol{\beta}$ in terms of the k -vector $\boldsymbol{\kappa}$ with $\boldsymbol{\xi}$ in (3.7) determined as $\xi_i = \kappa_i / \sum_{j=1}^k \kappa_j$, $i = 1, \dots, k$ (supplementary materials, Schofield and Barker, 2016, Section 1). The vector $\boldsymbol{\kappa}$ has only $k - 1$ free parameters as $\kappa_1 \equiv 1$. It follows that the statistics $\mathbf{u}|n$ are S-ancillary for \mathbf{p} and $\boldsymbol{\phi}$ in the presence of $\boldsymbol{\kappa}$ as we have reparameterized $(\mathbf{p}, \boldsymbol{\phi}, \boldsymbol{\beta})$ to $(\mathbf{p}, \boldsymbol{\phi}, \boldsymbol{\kappa})$ such that f_1 in (3.7) depends only on $\boldsymbol{\kappa}$. The MLEs for \mathbf{p} and $\boldsymbol{\phi}$ from the model of Link and Barker (2005) are identical to those from a CJS model.

Less clear is whether the MLEs for \mathbf{p} , $\boldsymbol{\phi}$ or $\boldsymbol{\kappa}$ are affected by conditioning on n . The term $\pi = \sum_{j=1}^k \psi_j p_j$ in (3.6) is not a free parameter. It can be seen from the recursion relationships in the supplementary materials, Schofield and Barker (2016), Section 1, that π is a complicated function of \mathbf{p} , $\boldsymbol{\phi}$ and $\boldsymbol{\kappa}$, so that n is not S-ancillary for \mathbf{p} , $\boldsymbol{\phi}$ and $\boldsymbol{\kappa}$. The MLEs from the full CMSA model will differ from the MLEs for models that condition on n or \mathbf{u} . However, as before, we can show that n is Bayes-ancillary for the parameters \mathbf{p} , $\boldsymbol{\phi}$ and $\boldsymbol{\kappa}$ in the presence of the prior distribution $f(\nu) \propto \nu^{-1}$. This means that conditioning on n (or \mathbf{u}) in place of using the full CMSA model can be justified with respect to use of a marginal likelihood for \mathbf{p} , $\boldsymbol{\phi}$ and $\boldsymbol{\kappa}$ where we have marginalized across ν with use of a scale prior.

As with the JS model, we can examine the effect of treating the random variable ν as a parameter as was

done by Schwarz and Arnason (1996). If we assume that v has a Poisson distribution with rate parameter μ_v , then marginalizing over v implies that the corresponding model for n is a Poisson model with rate parameter $\mu_v \pi$,

$$f_0(n|\mu_v, \mathbf{p}, \boldsymbol{\phi}, \boldsymbol{\beta}) = \frac{e^{-\pi\mu_v} (\pi\mu_v)^n}{n!},$$

in place of the binomial model given in (3.6). Assuming this model, the statistic n is S-ancillary for $\boldsymbol{\phi}$, \mathbf{p} and $\boldsymbol{\kappa}$ in the presence of μ_v . This implies that treating v as a Poisson random variable leads to identical MLEs for \mathbf{p} , $\boldsymbol{\phi}$ and $\boldsymbol{\kappa}$ to those from a model than conditions on n , but different MLEs to those found from the full CMSA model. All results above for the CMSA family of models are confirmed numerically (see supplementary materials, Schofield and Barker, 2016, Section 2).

3.3.4 Open-population generalization of Darroch's model B. Bayes-ancillarity also provides the necessary context for fully understanding the relationship between the JS model and the model of Pollock (1975) that generalized to open-populations model B of Darroch (1958). First consider the standard multinomial representation of a 2-sample closed-population capture-recapture study (model A of Darroch, 1958). We write the model as

$$\begin{aligned} f(\mathbf{x}^{\text{obs}}|v, \mathbf{p}) &= \frac{v!}{\prod_{\omega \in \Omega} x_{\omega}!} \prod_{\omega \in \Omega} \pi_{\omega}^{x_{\omega}} \\ (3.8) \quad &= \frac{\binom{n_1}{m_2} \binom{v-n_1}{n_2-m_2}}{\binom{v}{n_2}} \times \prod_{i=1}^2 \binom{v}{n_i} p_i^{n_i} (1-p_i)^{v-n_i} \\ &= f(m_2|\mathbf{n}, v) f(\mathbf{n}|\mathbf{p}, v), \end{aligned}$$

where $\Omega = \{00, 01, 10, 11\}$ and $n_1 = x_{11} + x_{10}$, $n_2 = x_{11} + x_{01}$, $m_2 = x_{11}$.

For inference about v , we can use $f(m_2|\mathbf{n}, v)$ as a partial likelihood for v , ignoring the information contributed by $f(\mathbf{n}|\mathbf{p}, v)$. This corresponds to model B of Darroch (1958); the MLE for this model is the well-known Lincoln-Petersen estimator

$$\hat{N} = \frac{n_1 n_2}{m_2}.$$

The statistics \mathbf{n} are not S-ancillary for v , but they are Bayes-ancillary in the presence of independent priors $f(p_i) \propto p_i^{-1}$ (Severini, 1995). Darroch (1958) argued this partial likelihood could be justified when the n_i are specified in advance, or “we could say that (model B)

is likely to be appropriate when the main limiting factor on sample size is the trouble involved in marking animals.” Here we have another justification, which is that use of the partial likelihood $f(m_2|\mathbf{n}, v)$ is justified if we adopt the prior distribution $f(\mathbf{p}) \propto \prod_i p_i^{-1}$ and use a marginal likelihood for inference concerning v . This interpretation extends to the open-population generalization of Darroch's model B given by Pollock (1975), with this model also conditioning on the sample sizes n_i .

3.4 Sufficiency and Goodness of Fit

Starting with the raw capture histories \mathbf{x}^{obs} and deriving the CMSA likelihood conveys another advantage. We can use factorizations of the likelihood that are functions of the data and sufficient statistics alone to determine the adequacy of the assumed model. For example, recall that in the model of Seber (1962) there was a term $f(\mathbf{x}^{\text{obs}}|\mathbf{m})$ that was ignored (since it is unnecessary for the likelihood). Similarly, we can further factorize the CJS model (2.2) as

$$f_3(\mathbf{x}^{\text{obs}}|\mathbf{R}, \mathbf{p}, \boldsymbol{\phi}) = f(\mathbf{x}^{\text{obs}}|\mathbf{m}, \mathbf{r}) f(\mathbf{m}, \mathbf{r}|\mathbf{R}, \mathbf{p}, \boldsymbol{\phi}).$$

The statistics \mathbf{m} and \mathbf{r} are sufficient statistics for \mathbf{p} and $\boldsymbol{\phi}$. The term $f(\mathbf{x}^{\text{obs}}|\mathbf{m}, \mathbf{r})$ is a product of multiple hypergeometric distributions and leads to various tests of model adequacy as described by Pollock, Hines and Nichols (1985).

4. LATENT VARIABLES

4.1 The Model of Jolly (1965)

A feature of Jolly (1965) is the use of latent variables to provide a state-space description of the problem, separating the capture and mortality components of the model. We do not believe this is well known, with the state-space relationship often attributed to Dupuis (1995), Gimenez et al. (2007) or Royle (2008). The ability of Jolly (1965) to separate capture and mortality is due to his inclusion of the latent variable M_{ij} , the true number of marked individuals last seen in occasion i still alive at the time of occasion j . The observation process of Jolly's state-space model separately considers marked and unmarked individuals,

$$f(u_j|U_j) = \binom{U_j}{u_j} p_j^{u_j} (1-p_j)^{U_j-u_j},$$

$$j = 1, \dots, k,$$

$$f(m_{ij}|M_{ij}) = \binom{M_{ij}}{m_{ij}} p_j^{m_{ij}} (1-p_j)^{M_{ij}-m_{ij}},$$

$$i = 1, \dots, k-1, j = i+1, \dots, k.$$

The state process separately models latent variables that describe the number of surviving (i) unmarked individuals, (ii) previously marked individuals (not caught in the current sample), and (iii) individuals caught in the current sample:

$$f(U_{j+1}|u_j, U_j, B_j) = \binom{U_j - u_j}{U_{j+1} - B_j} \phi_j^{U_{j+1} - B_j} (1 - \phi_j)^{U_j - u_j - U_{j+1} + B_j},$$

$$j = 1, \dots, k-1,$$

$$f(M_{ij+1}|m_{ij}, M_{ij}) = \binom{M_{ij} - m_{ij}}{M_{ij+1}} \phi_j^{M_{ij+1}} (1 - \phi_j)^{M_{ij} - m_{ij} - M_{ij+1}},$$

$$i = 1, \dots, k-2,$$

$$j = i+1, \dots, k-1,$$

$$f(M_{jj+1}|R_j) = \binom{R_j}{M_{jj+1}} \phi_j^{M_{jj+1}} (1 - \phi_j)^{R_j - M_{jj+1}},$$

$$j = 1, \dots, k-1.$$

As described in Section 2.5, Jolly treated M_{ij} , U_j and B_j as parameters to enable specification of a likelihood. Recent advances in statistical methodology now allow us to consider model fitting with computational approaches that account for the inclusion of latent variables. This is often referred to as using a complete data likelihood (CDL) and leads to the idea of modeling in terms of “the data we wished we had” (Draper, 1995). One such computational approach is Markov chain Monte Carlo (MCMC), where Monte Carlo integration is used to integrate over any latent variables specified in the model (Tanner and Wong, 1987). We provide JAGS (Plummer, 2003) code for fitting Jolly’s model above using MCMC. The model statement is available in the supplementary materials, Schofield and Barker (2016), Section 3, and a working example using the Dipper dataset of Lebreton et al. (1992) is available at http://www.stats.otago.ac.nz/webdata/resources/matthew_schofield/.

Two possible criticisms of the latent variable model above are as follows:

1. Many of the variables are of little interest. For example, we care little for M_{15} , as it depends on the sampling effort in occasions 1–4 as much as it does on the number of individuals alive in occasion 5.

2. A conditional probability model for B_j is not included. While we can predict the unknown number of births B_j , the dynamics of this process have not been described and modeled.

The latent variable model of Schofield, Barker and MacKenzie (2009, page 381) overcomes these two criticisms. Instead of modeling in terms of the sufficient statistics m_{ij} , the modeling is done in terms of the minimal sufficient statistics and is specified in terms of M_j , the total number of marked individuals alive at sampling occasion j . We note that this is only a modest improvement over the model of Jolly (1965), as it is straightforward in the latter to define $M_j = \sum_{i=1}^{j-1} M_{ij}$. The model of Schofield, Barker and MacKenzie (2009) then uses $N_j = M_j + U_j$ to specify a Poisson model for B_j in terms of a per-capita birth rate f_j ,

$$f(B_j) = \frac{e^{-f_j N_j} (f_j N_j)^{B_j}}{B_j!}, \quad j = 1, \dots, k-1.$$

Such an approach is related to those of Pradel (1996) and Link and Barker (2005), who also model the birth process in terms of a per-capita birth rate. A difference between those approaches and that specified here is that the birth rate is defined in terms of the latent population size N_j , whereas Pradel (1996) and Link and Barker (2005) derive the birth rate in terms of an expectation of N_j .

4.2 Latent Variables, Missing Data and the CMSA Model

As far as we are aware, the first instance of model fitting involving latent variables in a capture–recapture model is given by Dupuis (1995) in the context of the Arnason–Schwarz model (Brownie et al., 1993; Schwarz, Schweigert and Arnason, 1993) used to model animal movements. While this paper laid a solid theoretical foundation for future work, it was not until Dupuis and Schwarz (2007) and Schofield and Barker (2008) that corresponding versions of the CMSA model were implemented.

Schofield and Barker (2008) express the CMSA model in terms of a complete capture history and individual-specific times of birth and death, denoted b_i and d_i , for $i = 1, \dots, \nu$ respectively. From \mathbf{x}^{obs} we have partial information on these quantities. For example, we know that b_i , the time period in which individual i first appeared in a population, must have occurred at or before the sample in which that individual first appeared in the capture record. Similarly, d_i , the period in which the individual left the population, must have

occurred after the last sample in which the individual was captured.

Traditionally, when we construct the likelihood, we sum over all possible values for these partially observed quantities. The parameter ξ_j in the CMSA model represents $\Pr(\text{individual is first seen at } j)$ and is computed by summing over the probabilities of mutually exclusive outcomes. Similarly, the parameter χ_j represents $\Pr(\text{individual is not seen again after } j)$. Instead of using ξ and χ to marginalize over unobserved events, we can complete the data and use MCMC to integrate over the latent variables as described above.

If we suppose that the intervals of birth and death, \mathbf{b} and \mathbf{d} , are fully observed for each individual, the modeling is straightforward. Any information contained by the capture histories is irrelevant (under mild conditions it is S-ancillary for fecundity and survival rates). As such, we could directly model entry and exit rates using so-called known fate models (White and Burnham, 1999). Why then, if we include \mathbf{b} and \mathbf{d} into the model as latent variables, can we not treat the data as partially observed known fate data and ignore the model for the capture histories \mathbf{x}^{obs} ?

The answer is found by considering the latent variables as missing data (Dupuis, 1995; Schofield and Barker, 2008). Rubin (1976) and Little and Rubin (2002) define three classes of missing data: (i) missing completely at random (MCAR), (ii) missing at random (MAR) and (iii) missing not at random (MNAR). The important distinction between these classes is that if the missing data can be assumed MCAR or MAR, then the process by which the data went missing can be ignored in the statistical model, assuming that the parameters governing this process and the distribution of interest are distinct. In effect, the missing data process is S-ancillary to the unknown quantities of interest in the model. However, if the missing data are MNAR, then the process by which the data have gone missing contains information about the quantities of interest and should be included in the model.

For capture–recapture, the missing data in \mathbf{b} and \mathbf{d} are MNAR. As the process by which the data have gone missing is determined from the observation process, we need to include a model for \mathbf{x}^{obs} . The CDL offers a convenient factorization; we can specify a model for \mathbf{b} and $\mathbf{d}|\mathbf{b}$ along with a model for $\mathbf{x}^{\text{obs}}|\mathbf{b}, \mathbf{d}$,

$$(4.1) \quad \begin{aligned} f(\mathbf{x}^{\text{obs}}, \mathbf{b}, \mathbf{d}|\mathbf{p}, \phi, \beta) \\ = f(\mathbf{x}^{\text{obs}}|\mathbf{b}, \mathbf{d}, \mathbf{p})f(\mathbf{d}|\mathbf{b}, \phi)f(\mathbf{b}|\beta). \end{aligned}$$

Schofield and Barker (2009) show how the CDL allows for construction of the “mother-of-all-models.”

This idea, first suggested by Barker and White (2004), is that many of the distinct capture–recapture models are related to each other in terms of common likelihood components. As mentioned above, the use of a CDL shows the relationship between capture–recapture and other well-studied statistical approaches, such as state-space modeling and hidden Markov models (Langrock and King, 2013).

The simplest illustration of this approach is for the CJS model. If we define the occasion on which an individual was first released as t_i , the CDL is given by

$$f(\mathbf{x}^{\text{obs}}, \mathbf{d}|\mathbf{t}, \mathbf{p}, \phi) = f(\mathbf{x}^{\text{obs}}|\mathbf{t}, \mathbf{d}, \mathbf{p})f(\mathbf{d}|\mathbf{t}, \phi),$$

where

$$\begin{aligned} f(\mathbf{x}^{\text{obs}}|\mathbf{t}, \mathbf{d}, \mathbf{p}) &= \prod_{i=1}^n \prod_{j=t_i+1}^{d_i} p_j^{x_{ij}} (1-p_j)^{1-x_{ij}}, \\ f(\mathbf{d}|\phi) &= \prod_{i=1}^n \prod_{j=t_i}^k \theta_{ij}^{\mathbf{1}_{d_i=j}}, \end{aligned}$$

where $\mathbf{1}_E$ is the indicator for event E .

To fit these models, we require computational approaches that allow for latent variables. The method predominantly used in the literature is MCMC. When using MCMC methods, the latent variables d_i are included in the model and treated as another unknown quantity, like p_1 or ϕ_3 . If we specify independent and conditionally conjugate beta priors for each \mathbf{p} and ϕ , the model can be fitted using Gibbs sampling (Link and Barker, 2010, Section 11.3.1). The full conditional distribution for drawing values of d_i is a categorical distribution with support $\{l_i, l_i+1, \dots, k\}$, where l_i is the index of the last sample in which individual i was recorded.

An alternative to specifying the CMSA model in terms of \mathbf{b} and \mathbf{d} is to include this information in the $v \times k$ availability matrix \mathbf{a} , where the value $a_{ij} = 1$ if individual i is known to have been available for capture in sample j , $a_{ij} = 0$ if individual i is known to have been unavailable for capture in sample j and $a_{ij} = \text{NA}$ otherwise. Models specified in terms of \mathbf{a} are often used interchangeably with those in terms of \mathbf{b} and \mathbf{d} . However, it is likely that these different model specifications will lead to different MCMC algorithms. Our experience is that models written using \mathbf{a} make model specification in default programs such as BUGS (Lunn et al., 2009) and JAGS easier, but can mix poorly when compared to those specified in terms of \mathbf{d} . This is because the direct draw for d_i allows the Markov chain to move more quickly through the state space and converge more rapidly than corresponding algorithms for

a. Moreover, we believe that considering the model in terms of **d** opens additional possibilities in terms of improved algorithms in certain situations, such as using Gibbs multinomial probit regression (Imai and van Dyk, 2005; Jiao and van Dyk, 2015).

5. DISCUSSION

The distinction between many of the early contributions to capture–recapture is in how they treated random variables. We have focused on two particular aspects. The first involves the observed random variables that are treated as fixed. We have explored the impact that various choices have upon inference and have used ancillarity arguments to explain differences in the MLEs. For example, the MLEs for survival probabilities of the CJS model will differ from either the JS or CMSA model.

The second aspect is how unobserved random variables are accounted for when model fitting. A common feature of Jolly (1965) and Seber (1965) was that unobserved variables could be treated as parameters (and conditioned on) to overcome the difficulties encountered by Darroch (1959). A difference is that Jolly (1965) also included additional latent variables (M_{ij}) that allowed him to formulate what we now refer to as a state-space model. He also treated these as parameters, while the model of Seber (1965) marginalizes across these variables.

These differences are important and have only become more relevant in recent years. The explosion of state-space and latent variable models are following in the footsteps of Jolly (1965) and his model development. Often there are many possible model variants for a particular problem, distinctions that can be attributed to differences in the information that is being conditioned on. Notions of ancillarity can help us to better understand these differences.

Although this volume is a celebration of the work of Richard Cormack, George Jolly and George Seber, we believe it fitting to finish with an acknowledgement of the pioneering work of John Darroch. Darroch's name does not feature in either the Cormack–Jolly–Seber model or the Jolly–Seber model, but his contributions proved to be the foundation for modern analysis of capture–recapture data based on full probability models. Darroch got tantalizingly close to a complete solution in his 1959 paper with the stumbling block his attempt at a parameterization involving N . This was solved by Seber (1962, 1965) as well as Jolly (1965), the beginnings of a proliferation of capture–recapture models over the subsequent 50 years.

Although the early focus was on abundance estimation, development after the 1960s largely focused on extensions of the CJS model. Conditioning on the numbers released obviates the need to consider how marked and unmarked individuals differ, although, as noted by Cormack (1964), the marked individuals have to represent the study population in critical aspects for results to be generalizable. In a sense, the CMSA likelihood, and complete data implementations such as Dupuis and Schwarz (2007) and Schofield and Barker (2008), represents a completion of the work begun by Darroch in the late 1950s. They incorporate latent quantities such as abundances N_i and births B_i as well as parameters such as ϕ and β , providing an approach for the comprehensive study of population dynamics.

SUPPLEMENTARY MATERIAL

Supplementary materials for 50-year-old curiosities: ancillarity and inference in capture–recapture models (DOI: [10.1214/15-STS550SUPP](https://doi.org/10.1214/15-STS550SUPP); .pdf). We provide additional details about:

- The parameterization of the CMSA model that is used to determine the ancillarity results in Section 3.3.3.
- The numerical methods to confirm the theoretical results given in Section 3.3.
- A JAGS specification of the latent variable state-space model Jolly outlined in his 1965 paper.

REFERENCES

- BARKER, R. J. (1997). Joint modeling of live-recapture, tag-resight, and tag-recovery data. *Biometrics* **53** 666–677.
- BARKER, R. J. and WHITE, G. C. (2004). Towards the mother-of-all-models: Customised construction of the mark-recapture likelihood function. *Animal Biodiversity and Conservation* **27** 177–185.
- BARKER, R. J., SCHOFIELD, M. R., WRIGHT, J. A., FRANTZ, A. C. and STEVENS, C. (2014). Closed-population capture–recapture modeling of samples drawn one at a time. *Biometrics* **70** 775–782. [MR3295738](#)
- BARNDORFF-NIELSEN, O. (1976). Nonformation. *Biometrika* **63** 567–571. [MR0494594](#)
- BERGER, J. O., LISEO, B. and WOLPERT, R. L. (1999). Integrated likelihood methods for eliminating nuisance parameters. *Statist. Sci.* **14** 1–28. [MR1702200](#)
- BONNER, S. J. and SCHWARZ, C. J. (2006). An extension of the Cormack–Jolly–Seber model for continuous covariates with application to *Microtus pennsylvanicus*. *Biometrics* **62** 142–149, 317. [MR2226567](#)
- BOX, J. F. (1978). *R. A. Fisher: The Life of a Scientist*. Wiley, New York. [MR0500579](#)

- BROWNIE, C., ANDERSON, D. R., BURNHAM, K. P. and ROBINSON, D. A. (1976). Statistical inference from band recovery data—A handbook. Technical report, U.S. Department of the Interior, Fish and Wildlife Service Resource Publication No. 131, Washington, DC.
- BROWNIE, C., ANDERSON, D. R., BURNHAM, K. P. and ROBINSON, D. A. (1985). Statistical inference from band recovery data—A handbook. Technical report, U.S. Department of the Interior, Fish and Wildlife Service Resource Publication No. 156, Washington, DC.
- BROWNIE, C., HINES, J. E., NICHOLS, J. D., POLLOCK, K. H. and HESTBECK, J. B. (1993). Capture–recapture studies for multiple strata including non-Markovian transitions. *Biometrics* **49** 1173–1187.
- BURNHAM, K. P. (1991). On a unified theory for release-resampling of animal populations. In *Proceedings 1990 Taipei Symposium in Statistics* (M. T. Chao and P. E. Cheng, eds.) 11–35. Institute of Statistical Science, Academia Sinica, Taipei, Taiwan.
- BURNHAM, K. P. (1993). *A Theory for Combined Analysis of Ring Recovery and Recapture Data*. Birkhäuser, Basel.
- CORMACK, R. M. (1964). Estimates of survival from the sighting of marked animals. *Biometrika* **51** 429–438.
- CORMACK, R. M. (1989). Log-linear models for capture–recapture. *Biometrics* **45** 395–413.
- CORMACK, R. M. (1992). Interval estimation for mark-recapture studies of closed populations. *Biometrics* **48** 567–576. [MR1173495](#)
- COX, D. R. (1975). Partial likelihood. *Biometrika* **62** 269–276. [MR0400509](#)
- CROSBIE, S. F. (1979). The mathematical modelling of capture-mark-recapture experiments on animal populations. Ph.D. thesis, University of Otago, Dunedin, New Zealand.
- CROSBIE, S. F. and MANLY, B. F. J. (1985). Parsimonious modelling of capture-mark-recapture studies. *Biometrics* **41** 385–398.
- DARROCH, J. N. (1958). The multiple-recapture census. I. Estimation of a closed population. *Biometrika* **45** 343–359. [MR0119360](#)
- DARROCH, J. N. (1959). The multiple-recapture census. II. Estimation when there is immigration or death. *Biometrika* **46** 336–351. [MR0119361](#)
- DRAPER, D. (1995). Inference and hierarchical modeling in the social sciences. *J. Educ. Behav. Stat.* **20** 115–147.
- DUPUIS, J. A. (1995). Bayesian estimation of movement and survival probabilities from capture–recapture data. *Biometrika* **82** 761–772. [MR1380813](#)
- DUPUIS, J. A. and SCHWARZ, C. J. (2007). A Bayesian approach to the multistate Jolly–Seber capture–recapture model. *Biometrics* **63** 1015–1022, 1309. [MR2414577](#)
- FISHER, R. A. (1925). Theory of statistical estimation. In *Mathematical Proceedings of the Cambridge Philosophical Society* **22** 700–725. Cambridge Univ Press, Cambridge.
- FISHER, R. A. (1934). Two new properties of mathematical likelihood. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **144** 285–307.
- FISHER, R. A. and FORD, E. B. (1947). The spread of a gene in natural conditions in a colony of the moth *Panaxia dominula* L. *Heredity* **1** 143–174.
- GHOSH, M., REID, N. and FRASER, D. A. S. (2010). Ancillary statistics: A review. *Statist. Sinica* **20** 1309–1332. [MR2777327](#)
- GIMENEZ, O., ROSSI, V., CHOQUET, R., DEHAIS, C., DORIS, B., VARELLA, H., VILA, J. P. and PRADEL, R. (2007). State-space modelling of data on marked individuals. *Ecological Modelling* **206** 431–438.
- IMAI, K. and VAN DYK, D. A. (2005). A Bayesian analysis of the multinomial probit model using marginal data augmentation. *J. Econometrics* **124** 311–334. [MR2125369](#)
- JACKSON, C. H. N. (1939). The analysis of an animal population. *Journal of Animal Ecology* **8** 238–246.
- JIAO, X. and VAN DYK, D. A. (2015). A corrected and more efficient suite of MCMC samplers for the multinomial probit model. Preprint. Available at [arXiv:1504.07823](#).
- JOLLY, G. M. (1963). Estimates of population parameters from multiple recapture data with both death and dilution—Deterministic model. *Biometrika* **50** 113–128. [MR0157460](#)
- JOLLY, G. M. (1965). Explicit estimates from capture–recapture data with both death and immigration-stochastic model. *Biometrika* **52** 225–247. [MR0210227](#)
- KENDALL, W. L., POLLOCK, K. H. and BROWNIE, C. (1995). A likelihood-based approach to capture–recapture estimations of demographic parameters under the robust design. *Biometrics* **51** 293–308.
- KING, R., BROOKS, S. P., MORGAN, B. J. T. and COULSON, T. (2006). Factors influencing Soay sheep survival: A Bayesian analysis. *Biometrics* **62** 211–220, 319. [MR2226575](#)
- LANGROCK, R. and KING, R. (2013). Maximum likelihood estimation of mark-recapture-recovery models in the presence of continuous covariates. *Ann. Appl. Stat.* **7** 1709–1732. [MR3127965](#)
- LEBRETON, J. D., BURNHAM, K. P., CLOBERT, J. and ANDERSON, D. R. (1992). Modelling survival and testing biological hypotheses using marked animals: A unified approach with case studies. *Ecological Monographs* **62** 67–118.
- LESLIE, P. H., CHITTY, D. and CHITTY, H. (1953). The estimation of population parameters from data obtained by means of the capture–recapture method: III. An example of the practical applications of the method. *Biometrika* **40** 137–169.
- LINK, W. A. and BARKER, R. J. (2005). Modeling association among demographic parameters in analysis of open population capture–recapture data. *Biometrics* **61** 46–54. [MR2129200](#)
- LINK, W. A. and BARKER, R. J. (2010). *Bayesian Inference with Ecological Applications*. Academic Press, London.
- LITTLE, R. J. A. and RUBIN, D. B. (2002). *Statistical Analysis with Missing Data*, 2nd ed. Wiley, Hoboken, NJ. [MR1925014](#)
- LUNN, D., SPIEGELHALTER, D., THOMAS, A. and BEST, N. (2009). The BUGS project: Evolution, critique and future directions. *Stat. Med.* **28** 3049–3067. [MR2750401](#)
- NICHOLS, J. D. (2016). And the first one now will later be last: Time-reversal in Cormack–Jolly–Seber models. *Statist. Sci.* **31** 175–190.
- PLUMMER, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*. Vienna, Austria.
- POLLOCK, K. H. (1975). A K -sample tag-recapture model allowing for unequal survival and catchability. *Biometrika* **62** 577–583. [MR0400517](#)

- POLLOCK, K. H. (1982). A capture–recapture design robust to unequal probability of capture. *Journal of Wildlife Management* **46** 752–757.
- POLLOCK, K. H., HINES, J. E. and NICHOLS, J. D. (1985). Goodness-of-fit tests for open capture–recapture models. *Biometrics* **41** 399–410.
- POLLOCK, K. H., NICHOLS, J. D., BROWNIE, C. and HINES, J. E. (1990). Statistical inference for capture–recapture experiments. *Wildlife Monographs* **107**.
- PRADEL, R. (1996). Utilization of capture–mark–recapture for the study of recruitment and population growth rate. *Biometrics* **52** 703–709. [MR1395003](#)
- ROYLE, J. A. (2008). Modeling individual effects in the Cormack–Jolly–Seber model: A state-space formulation. *Biometrics* **64** 364–370, 664. [MR2432405](#)
- RUBIN, D. B. (1976). Inference and missing data. *Biometrika* **63** 581–592. [MR0455196](#)
- SANATHANAN, L. (1972). Estimating the size of a multinomial population. *Ann. Math. Stat.* **43** 142–152. [MR0298815](#)
- SANDLAND, R. L. and CORMACK, R. M. (1984). Statistical inference for Poisson and multinomial models for capture–recapture experiments. *Biometrika* **71** 27–33. [MR0738322](#)
- SCHOFIELD, M. R. and BARKER, R. J. (2008). A unified capture–recapture framework. *J. Agric. Biol. Environ. Stat.* **13** 458–477. [MR2590940](#)
- SCHOFIELD, M. R. and BARKER, R. J. (2009). A further step toward the mother-of-all-models: Flexibility and functionality in the modeling of capture–recapture data. In *Modeling Demographic Processes in Marked Populations* (D. L. Thomson, E. G. Cooch and M. J. Conroy, eds.). *Environmental and Ecological Statistics* **3** 677–689. Springer, New York.
- SCHOFIELD, M. R. and BARKER, R. J. (2016). Supplement to “50-Year-Old Curiosities: Ancillarity and Inference in Capture–Recapture Models.” DOI:[10.1214/16-STS550SUPP](#).
- SCHOFIELD, M. R., BARKER, R. J. and MACKENZIE, D. I. (2009). Flexible hierarchical mark–recapture modeling for open populations using WinBUGS. *Environ. Ecol. Stat.* **16** 369–387. [MR2749846](#)
- SCHWARZ, C. J. and ARNASON, A. N. (1996). A general methodology for the analysis of capture–recapture experiments in open populations. *Biometrics* **52** 860–873. [MR1411736](#)
- SCHWARZ, C. J., SCHWEIGERT, J. F. and ARNASON, A. N. (1993). Estimating migration rates using tag–recovery data. *Biometrics* **49** 177–193.
- SCHWARZ, C. J. and SEBER, G. A. F. (1999). A review of estimating animal abundance III. *Statist. Sci.* **14** 427–456.
- SEBER, G. A. F. (1962). The multi-sample single recapture census. *Biometrika* **49** 339–350. [MR0157458](#)
- SEBER, G. A. F. (1965). A note on the multiple-recapture census. *Biometrika* **52** 249–259. [MR0210228](#)
- SEBER, G. A. F. (1970). Estimating time-specific survival and reporting rates for adult birds from band returns. *Biometrika* **57** 313–318.
- SEBER, G. A. F. (1982). *The Estimation of Animal Abundance and Related Parameters*, 2nd ed. Charles Griffin and Company, High Wycombe, England. [MR0686755](#)
- SEBER, G. A. F. (1986). A review of estimating animal abundance. *Biometrics* **42** 267–292.
- SEBER, G. A. F. (1992). A review of estimating animal abundance II. *Int. Stat. Rev.* **60** 129–166.
- SEVERINI, T. A. (1995). Information and conditional inference. *J. Amer. Statist. Assoc.* **90** 1341–1346. [MR1379476](#)
- TANNER, M. A. and WONG, W. H. (1987). The calculation of posterior distributions by data augmentation. *J. Amer. Statist. Assoc.* **82** 528–550. [MR0898357](#)
- WHITE, G. C. and BURNHAM, K. P. (1999). Program MARK: Survival estimation from populations of marked animals. *Bird Study* **46** S120–S139.
- WILLIAMS, B. K., NICHOLS, J. D. and CONROY, M. J. (2002). *Analysis and Management of Animal Populations*. Academic Press, San Diego, CA.