Hierarchical mark–recapture models: a framework for inference about demographic processes

W. A. Link & R. J. Barker

Abstract

Hierarchial mark–recapture models: a framework for inference about demographic processes.— The development of sophisticated mark–recapture models over the last four decades has provided fundamental tools for the study of wildlife populations, allowing reliable inference about population sizes and demographic rates based on clearly formulated models for the sampling processes. Mark–recapture models are now routinely described by large numbers of parameters. These large models provide the next challenge to wildlife modelers: the extraction of signal from noise in large collections of parameters. Pattern among parameters can be described by strong, deterministic relations (as in ultrastructural models) but is more flexibly and credibly modeled using weaker, stochastic relations. Trend in survival rates is not likely to be manifest by a sequence of values falling precisely on a given parametric curve; rather, if we could somehow know the true values, we might anticipate a regression relation between parameters and explanatory variables, in which true value equals signal plus noise. Hierarchical models provide a useful framework for inference about collections of related parameters. Instead of regarding parameters as fixed but unknown quantities, we regard them as realizations of stochastic processes governed by hyperparameters. Inference about demographic processes is based on investigation of these hyperparameters. We advocate the Bayesian paradigm as a natural, mathematically and scientifically sound basis for inference about hierarchical models. We describe analysis of capture–recapture data from an open population based on hierarchical extensions of the Cormack–Jolly–Seber model. In addition to recaptures of marked animals, we model first captures of animals and losses on capture, and are thus able to estimate survival probabilities φ (i.e., the complement of death or permanent emigration) and per capita growth rates f (i.e., the sum of recruitment and immigration rates). Covariation in these rates, a feature of demographic interest, is explicitly described in the model.

Key words: Bayesian hierarchical analysis, Capture–recapture, Demographic analysis, Jolly–Seber Model, Open population estimation.

Resumen

Modelos jerárquicos de marcaje–recaptura: un marco para la inferencia de procesos demográficos.— El desarrollo de sofisticados modelos de marcaje–recaptura a lo largo de las últimas cuatro décadas ha proporcionado herramientas fundamentales para el estudio de poblaciones de fauna silvestre, lo que ha permitido inferir con fiabilidad los tamaños poblacionales y las tasas demográficas a partir de modelos claramente formulados para procesos estocásticos. En la actualidad, los modelos de marcaje–recaptura se describen de forma rutinaria mediante una extensa serie de parámetros. Dichos modelos representan el siguiente reto al que deberán enfrentarse los modeladores de fauna silvestre: discriminar las señales del ruido en amplias series de parámetros. La pauta que encontramos en los parámetros puede describirse mediante sólidas relaciones deterministas (como en los modelos ultraestructurales), pero resulta más flexible y creíble si se modela utilizando relaciones estocásticas más débiles. No es probable que la tendencia en las tasas de supervivencia se manifieste por una secuencia de valores hallados concretamente en una curva paramétrica dada; por ello, si pudiéramos llegar a conocer los valores reales, podríamos prever una relación de regresión entre parámetros y variables explicativas, de forma que el valor verdadero

equivaldría a la señal más el ruido. Los modelos jerárquicos proporcionan un marco útil para la inferencia acerca de series de parámetros relacionados. Así, en lugar de interpretar los parámetros como cantidades fijas, pero desconocidas, los interpretamos como realizaciones de procesos estocásticos regidos por hiperparámetros. La inferencia acerca de los procesos demográficos se basa en la investigación de dichos hiperparámetros. Por este motivo, defendemos el paradigma bayesiano como una base natural, matemática y científicamente sólida para la inferencia acerca de modelos jerárquicos. En el presente estudio describimos el análisis de datos de captura–recaptura obtenidos a partir de una población abierta basada en ampliaciones jerárquicas del modelo de Cormack–Jolly–Seber. Además de las recapturas de animales marcados, también modelamos las primeras capturas de animales y de pérdidas durante la captura, lo que nos permitió estimar las probabilidades de supervivencia de φ (es decir, el complemento de la muerte o la emigración permanente) y las tasas de crecimiento per cápita f (es decir, la suma de las tasas de reclutamiento y de migración). En el modelo se describe explícitamente la covariación en estas tasas, que constituye una característica de interés demográfico.

Palabras clave: Análisis jerárquico bayesiano, Captura–recaptura, Análisis demográfico, Modelo de Jolly–Seber, Estimación de población abierta.

William A. Link, Patuxent Wildlife Research Center, Laurel, MD, 20708, U.S.A.– Richard J. Barker, Univ. of Otago, Dunedin, New Zealand.

Introduction

Observations of biological systems are variable. Mathematical models describing this variability incorporate two sources of variation: variation relating to the biological system itself (birth, death, and behavioral processes), and variation relating to the collection of data. Thus there are two categories of parameters in models for biological data. We will let θ denote the collection of demographic parameters describing biological processes, and η denote the collection of parameters governing collection of data; we will assume that these are unrelated, though this need not be the case.

The parameters comprising θ are the parameters of interest. Those comprising η are often referred to as "nuisance parameters". It is an appropriate name: their inclusion in the modeling effort is a necessary evil, and one which must be handled with care, because incorrect assumptions about η can lead to profound biases in estimation of θ.

The field of capture–recapture has developed over the last forty years with its first priority being on accounting for biologically irrelevant sources of variation in data. Increases in computational efficiency have allowed for analysis of larger data sets which not only include many nuisance parameters, but also many parameters of interest. The vector θ may have hundreds of components, including survival rates, recruitment rates, population sizes, movement parameters. We believe that the next priority in capture–recapture analysis should be the development of statistically sound methods for analysis, not of data, but of the parameters comprising θ.

In this paper, we describe hierarchical models useful in examining pattern in parameters, making note of various ad hoc methods that have been used for examining them, and arguing for the usefulness and appeal of the Bayesian paradigm in this context. We illustrate our discussion with an hierarchical extension of the Cormack–Jolly–Seber model which allows for efficient, statistically sound analysis of covariation among demographic rates.

Hierarchical models defined; *ad hoc* analytic methods

Suppose that we did not have to deal with nuisance parameters, indeed, that we did not need to estimate θ, but knew its components without error. What would we do? Why, we'd do some statistical analyses, regressing survival rates against time, examining whether recruitment rates were related to population size, and performing similar analyses. After all, the entire purpose of collecting biological data is to make inference about θ.

We would be mightily surprised if there were perfect, deterministic relations among the actual parameters. That is, if φ_t is a survival rate for time period $[t, t+1)$, we would not expect

$$\log(\varphi_t) = A + Bt \qquad (1)$$

though we might anticipate acceptable fit of a model with

$$\log(\varphi_t) = a + bt + \varepsilon_t \qquad (2)$$

where the values ε_t are a sample from a mean–zero normal distribution. The first of these specifies strong, deterministic relations among parameters; the survival rates are perfectly predictable, changing in lockstep fashion through time. The second specifies weaker, stochastic relations among parameters, general tendencies rather than predetermined patterns. The role of the error term ε_t is to account for sources of variation that are not explained by known covariates, but that nevertheless are an important source of variation to the survival probabilities; modeling thus, we assume that the collective effects of unknown covariates amount to uncorrelated random noise.

All statistical analyses, even nonparametric ones, begin with the specification of a family of distributions from which data have been drawn. Parametric analyses restrict the family to a general form, known except for certain unknown parameters, typically regarded (in the Frequentist paradigm) as "fixed but unknown constants". Hierarchical models treat these parameters as though they also were sampled from parametric families of distributions. Thus, *many* data are described by model specifications and *fewer* parameters; these parameters are described by further model specifications and even fewer parameters.

The simplest hierarchical models are familiar as random effects models. Indeed, the notion of finding and evaluating pattern among parameters is no new one. What is changing is the manner in which such models are fit to data. Over the years, the approach to investigating stochastic relations among demographic parameters has been one of two sorts (Link, 1999). First, we might fit an unconstrained model, treating components of θ as completely unrelated, then attempting to uncover the stochastic relations among parameters by examination of their estimates $\hat{\theta}$. Thus, for example, if we were interested in examining temporal pattern of change in survival rates, we might fit the model

$$\log(\hat{\varphi}_t) = a^* + b^*t + \varepsilon_t^* \qquad (3)$$

instead of (2), in the hope that inference about b^* in (3) could inform us about the parameter b in (2). This approach could be called a two–stage modeling approach: first estimate the parameters without specifying relations among them; then, look for those relations among the parameter estimates obtained.

The other commonly used approach to examining pattern among parameters has been ultrastructural modeling, in which the pattern among parameters is treated as deterministic. To examine temporal patterns of change in survival, we estimate the parameters φ_t subject to the constraint given by equation (1). Thus the parameter set is reduced from $\{\varphi_1, \varphi_2,...\}$ to $\{A, B\}$. We attempt to model $\{a, b\}$ using the pair $\{A, B\}$.

The two–stage and ultrastructural approach are approximate and unsatisfactory: in two–stage analyses, there may be problems of attenuation bias when random regressors are used. There may be inefficiencies, even gross ones, due to heterogeneity of variances among the estimators. Failure to adequately account for the sampling variability of parameter estimates, when using these as surrogates for actual parameter values, can be spectacularly misleading. Ultrastructural analyses often (but not inevitably) produce reasonable estimates, though usually with overstated precision. Various methods for improving ultrastructural and two–stage analyses have been developed, but may generally be said to only work asymptotically. In using them, we ought not to forget Le Cam's 7th principle: "If you need to use asymptotic arguments, do not forget to let your number of observations tend to infinity" (Le Cam, 1990).

Hierarchical models and Bayesian analysis

Instead of using ad hoc methods of analysis for hierarchical models, we recommend the use of Bayesian methods. The first advantage of these is the naturalness of their application to hierarchical models. All quantities in Bayesian inference are random variables; the only distinction being whether quantities are known or not. Thus the hierarchical modeling view of parameters as random variables is completely natural to Bayesian analysis.

Second, Bayesian analysis, properly conducted, requires no fussing over optimality criteria, choice of estimation technique, or asymptotics. There is nothing but a calculation, the calculation of a posterior distribution from a prior and a likelihood. All inference is based on features of the posterior distribution.

But what of the prior? Choice of the prior distribution is the classically trained statistician's Bayesian bugaboo. Choice of the prior is thought by some to introduce an irremediable and unacceptable subjectivity into the analysis. Attempts to define noninformative priors for objective analysis are dismissed by some critics on the grounds that the quality of noninformativeness is not transformation invariant. For instance, a noninformative prior for σ^2 is not a noninformative prior for σ. It is a marvelous thing that such critics will frequently choose the $(n-1)$ weighted variance estimator S^2 as an estimator of σ^2 because it is unbiased, giving no thought to the fact that S, the value they will actually use in discussion, is a biased estimator of σ. Unbiasedness, like noninformativeness of priors, is not transformation invariant. Lack of transformation invariance is a problem for Frequentists as well as for Bayesians.

The simple solution to the problem of choosing priors is to try several, and to see whether and how the choice influences posterior inference. In our view, this is a virtue rather than a vice of Bayesian inference. What is more, the Bernstein–Von Mises Theorem (also known as the Bayesian Central Limit

theorem) implies that, subject to minor constraints, the influence of the prior diminishes as the sample size increases, so that the choice will not matter if sample sizes are reasonably large.

Putting aside all of the overheated and overblown Bayesian/Frequentist polemics, we find that the differences of inference generally prove to be slight, provided we have adequate data. If we are willing to adopt the Bayesian paradigm, we may avail ourselves of the powerful computational techniques known as Markov chain Monte Carlo (MCMC) to examine features of the posterior distribution. And even if we simply cannot shake misgivings about the Bayesian paradigm, MCMC is still useful: by specifying uniform prior distributions, the posterior distribution is simply the scaled likelihood, and its mode the maximum likelihood estimator. We thus obtain a tool for formal examination of complex demographic structures in data sets; specification of stochastic relations among demographic parameters becomes part of data modeling, rather than something done after the fact using dubious ad hoc methods.

An analysis of association among demographic parameters

We illustrate the use of hierarchical models with an analysis of temporal relations between demographic parameters governing rates of population change. These are survival rates φ (i.e., the complement of death or permanent emigration) and per capita growth rates f (i.e., the sum of recruitment and immigration rates); the analysis will allow formal evaluation and testing of associations between these demographic parameters. We analyze the *Gonodontis* data set of Bishop et al. (1978), which has been evaluated in several subsequent papers developing methods for open–population survival analysis (Crosbie, 1979; Crosbie & Manly, 1985; Link & Barker, in press); the model we describe is an extension of the Cormack–Jolly–Seber (CJS) model (Cormack, 1964; Jolly, 1965; Seber, 1965).

The CJS model uses likelihood proportional to the joint distribution of sufficient statistics *r* and *m*, given statistics *u* and *R*. These statistics are vector–valued; R_i from *R* is the number of individuals released after the ith sampling occasion, and r_i from *r* is the number of these subsequently recaptured. Components of *m* and *u* are the numbers of marked and unmarked animals, respectively, at each of the sampling occasions. The conditional distribution corresponding to the CJS model can be written in self–explanatory notation as

$$[r, m \mid R, u] \qquad (4)$$

The dependence of this distribution on demographic and nuisance parameters is suppressed in this notation. We note however that the only demographic parameters are survival rates φ_i. Neither population size nor population growth rates are

included in the model. An extension of the CJS model is needed; we propose one subsequently.

Step one: obtaining a likelihood based on parameters of interest

The first step in conducting a Bayesian hierarchical analysis is the specification of a likelihood for the data in terms of the parameters of interest. This is generally straightforward, but can sometimes require some thought about reparameterizing and extending models.

The first attempt to extend the CJS model so as to allow examination of population change was the Jolly–Seber (JS) model. This extension had two components. The first, relatively minor change, is the addition of a parametric description of loss on capture, thus removing the conditioning on R, so that the likelihood used is proportional to

$$[r, m, R \mid u] \qquad (5)$$

Of greater importance is the modeling of first captures u_i as binomial with index U_i. Identifiability of U_i, the number of unmarked animals in the population just prior to the ith sampling occasion, is obtained by assuming that the nuisance parameters (detection probabilities) are the same for marked and unmarked animals. The JS model thus uses a likelihood proportional to the joint distribution

$$[r, m, R \mid u] \ [u \mid U] \qquad (6)$$

The population numbers of unmarked animals, U_i, are not of much interest *per se*, for investigating relations between survival and other parameters governing population growth rates. The model is not expressed in terms of quantities in which we are directly interested. If we want to use the JS model for our hierarchical investigations, we are forced to carry out a two–stage analysis, first obtaining estimates of demographic parameters (first, we estimate the population number of marked animals, and add this to the estimated number of unmarked animals to obtain an estimate of population size; more precisely, we should say, "predict" the population size, since population size is not part of the JS model; with the additional assumption that survival rates are the same for marked and unmarked animals, the resulting estimates of population change can be partitioned among changes due, on the one hand, to mortality, and to nonmortality sources, on the other), and then looking for associations among the estimates.

Instead, we consider an alternative extension of the CJS model, following work by Crosbie & Manly (1985), and later developed by Schwarz & Arnason (1996). Schwarz & Arnason describe the model in terms of a likelihood proportional to

$$[r, m, R \mid u] [u \mid N^*]$$

where N^* is the number of distinct animals available for capture on at least one of the sampling occasions. Their model includes the additional assumption that survival and detection rates are the same for marked and unmarked animals.

We demonstrate elsewhere that

$$[u \mid N^*] = [u \mid u.] [u. \mid N^*]$$

where $u. = \Sigma_i u_i$, and that there is very little information in the likelihood component proportional to $[u. \mid N^*]$ (Link & Barker, in press). We thus eliminate that from the likelihood, and base our analysis on

$$[r, m, R \mid u] [u \mid u.] \qquad (7)$$

It can be shown that the distribution $[u \mid u.]$ is multinomial with index $u.$, and cell probabilities ξ_i determined by detection and survival rates from the CJS model and $t - 1$ additional estimable parameters, namely $\rho_1, f_2, f_3, ..., f_{t-2}$ and ρ_t. Here, parameters f_i (slightly different from the growth rate parameters of Pradel et al., 1996; for details, see Link & Barker, in press) are per capita growth rates (recruitment plus immigration). Parameters ρ_1 and ρ_t are confounded combinations of demographic and nuisance parameters. Details on these and the functional form of ζ_i are given in the Appendix.

We use (7) as the basis for hierarchical analysis of relations between survival and growth rates rather than (5), because it is expressed entirely in terms of the demographic parameters of interest, and a clearly identified set of nuisance parameters.

We have

$$\boldsymbol{\theta} = \{\boldsymbol{\varphi}; \boldsymbol{f}\} = \{\varphi_1, \varphi_2, ..., \varphi_{t-2}; f_2, f_3, ..., f_{t-2}\}$$

and

$$\boldsymbol{\eta} = \{\rho_2, \rho_3, ..., \rho_{t-1}; v_2, v_3, ..., v_{t-1}; \rho_1, \rho_t, \lambda_{t-1}\}$$

here, t is the number of sampling occasions, ρ_i is the detection probability at sampling occasion i, v_i is the probability of successfully releasing an animal captured at sampling occasion i, and $\lambda_{t-1} = \varphi_{t-1} \rho_t$.

Step two: describing stochastic relations among parameters of interest

The next step in a hierarchical analysis is a description of stochastic relations among parameters of interest. From the Bayesian perspective, as we shall see subsequently, this amounts to a partial specification of the prior distributions of parameters.

Our goal is to examine stochastic relations between φ_i's and f_i's. Since $0 < \varphi_i < 1$ and $f_i > 0$, it is natural to transform the parameters in order to remove the range restrictions. We thus suppose that pairs

$$\theta_i = \{\text{logit} \ (\varphi_i) \ \log \ (f_i)\}$$

follow a bivariate normal distribution with mean vector μ and variance matrix Σ. The correlation parameter in the variance matrix is of primary interest, as determining the strength of association between φ_i's and f_i's . We will refer to $\Psi = \{\mu, \Sigma\}$ as the hyperparameters.

The process of specifying this part of the model is similar to the corresponding process for modeling stochastic relations among data. Ideally, there should be some basis in first principles for deciding on whether data are normally distributed, or whether a Poisson distribution, or Binomial is appropriate. Alternatively, distributional choices are often made on the basis of convenience. In observational studies, these choices are usually made after informal inspection of the data; similar informal evaluations of *parameters* might be based on the *ad hoc* methods described at the outset. In any case, the same sort of model checking used for evaluating distributional assumptions for *data* should be used to evaluate distributional assumptions about *parameters*.

Step three: selection of prior distributions

So far, we have suppressed in our notation the dependence of the likelihoods on the parameters. In order to complete the specification of a Bayesian hierarchical model, we make this dependence explicit, rewriting (7) as

$$[r, m, R \mid u; \theta, \eta \,] \, [u \mid u.; \theta, \eta]$$

All that remains for a fully Bayesian specification of the model, is a description of prior distributions for parameters and hyperparameters, i.e. $[\theta, \eta, \Psi]$. Here, we assume prior independence of the nuisance and demographic parameters, so that

$$[\theta, \eta, \Psi] = [\theta, \Psi] \, [\eta] = [\theta \mid \Psi] \, [\Psi] \, [\eta]$$

Our study of association among demographic parameters will be based on the partial specification of the prior structure, namely $[\theta \mid \Psi]$. The posterior distribution upon which we shall base inference is proportional to the joint distribution

$$[r, m, R \mid u; \theta, \eta \,] \, [u \mid u.; \theta, \eta] \, [\theta \mid \Psi] \, [\Psi] \, [\eta] \qquad (8)$$

All that remains is specification of the prior distributions $[\Psi]$ and $[\eta]$. We chose flat priors $[\eta] \propto c$, for the nuisance parameters.

We chose a Normal–Inverse Wishart prior distribution for Ψ. This is the distribution obtained by supposing that $[\mu, \Sigma]$ is a bivariate normal distribution with mean μ_0 and variance matrix 0, and that Σ/n_0 has the inverse Wishart distribution with parameters V and df. It is an appealing choice for a prior distribution for parameters of the multivariate normal, because the resulting posterior distributions for Ψ are easily calculated, also being members of the inverse Wishart family of distributions. The inverse Wishart distribution can be thought of
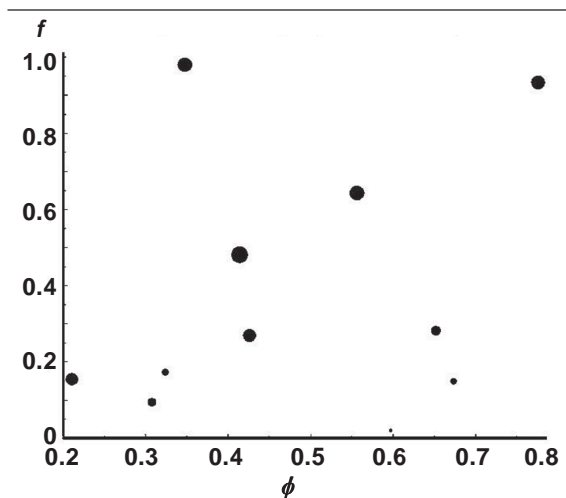

Fig. 1. Annual survival *vs.* nonsurvival growth rates (symbol size proportional to precision).

Fig. 1. Tasas de crecimiento de supervivencia anual respecto a no supervivencia (el tamaño de los símbolos es proporcional a la precisión).

as the distribution of

$$\Sigma = \left[\sum_{i=1}^{df} X_i X_i' \right]^{-1}$$

where X_i are independent and identically distributed bivariate normal random variables with variance matrix V^{-1}. Choice of the particular Normal–Inverse Wishart prior requires specification of μ_0, $n_0 \geq 0$, V, and $df \geq 2$.

For the *Gonodontis* data set $t = 17$. Recognizing that there are only $t - 3 = 14$ estimable pairs (φ_i, f_i) informing our inference about the covariation, we anticipated some sensitivity to the choice of parameters governing the choice of NIW priors for (μ, Σ), hence decided to repeat the analysis for four choices of prior. We set $n_0 = 0$ in all analyses, inducing vague priors on μ. Choice of parameters for the inverse Wishart distribution for Σ was guided by the observations that if $df \geq 1$,

1) The diagonal elements of Σ have inverse Gamma distributions:

$$\frac{1}{\sigma_i^2} \sim \Gamma\left(\frac{df - 1}{2}, \frac{V_{i,i}}{2} \right)$$

where $V_{i,i}$ is the ith diagonal element of V, so that σ_i^2 has the same distribution as $V_{i,i} / A$, where $A \sim \chi^2_{df-1}$, and

2) That given V is a diagonal matrix, the marginal distribution of the correlation parameter p is such that

$$(p + 1)/2 \sim \beta\left(\frac{df - 1}{2}, \frac{df - 1}{2} \right)$$

The four priors we considered were: (1) $df = 2$,

V = diag (1, 1); (2) *df* = 2, V = diag (2.5, 1.25); (3) *df* = 3, V = diag (1, 1); (4) *df* = 3, V = diag (5.0, 2.5).

The first two priors maximize the coefficient of variation of the diagonal elements of Σ, subject to the requirement that $[\Sigma]$ be proper; the third and fourth induce a uniform prior on the correlation parameter. The second and fourth fix the prior means of the precision (inverse variance) for logit survival rates and log birth rates at values of 0.40 and 0.80, respectively; these values were chosen as representing large levels of variability in demographic parameters. The first and third priors were chosen as representative of moderate levels of variability in demographic parameters.

Analysis of correlation in *Gonodontis* data

Figure 1 is a scatter plot of maximum likelihood estimates $\left(\hat{f}_i, \hat{\varphi}_i\right)$ based on the Pradel (1996) model, as implemented in program MARK (White & Burnham, 1999); note that the size of the plotting image reflects the relative precision of estimates. Three points, with imprecise estimates on the edge of the parameter space, were excluded.

The plot suggests a positive association between the vital rates. We wish to study the association between log(f) and logit(φ). However, uncovering this association by "doing statistics on statistics" is problematic. First, how are we to transform estimates $\hat{\varphi}$ = 1 to the logit scale? How are we to account for sampling variation, and covariation? Are the asymptotic variance estimates obtained from the estimated Fisher information matrix reliable? And supposing that all of the uncertainties raised by

these questions can be dismissed, what about computing a confidence interval for the estimated correlation parameter, $\hat{\rho}$? Are we confident that the estimator has a normal distribution, even though there are only 14 values (estimates, at that) from which it is calculated?

Far more satisfactory, to our mind, is the Bayesian analysis we have developed in the foregoing sections. We find relatively minor differences among the results based on four distinct priors; these are summarized in plots of the posterior distributions for p in figure 2, and on summaries of these distributions in table 1. The posterior probability that p > 0 is roughly 84% for all of the priors considered (0.843, 0.857, 0.816 and 0.843, for priors 1, 2, 3, and 4); the posterior odds are 5:1 in favor of a positive correlation.

Summary and comments

Associations among demographic parameters are very naturally modeled by treating the parameters as random variables, the associations arising because the parameters have been sampled from related distributions. Such hierarchical models posit the existence of weak, stochastic relations among parameters, rather than unrealistic deterministic relations. The Bayesian paradigm, in which *all* quantities are treated as stochastic, is particularly appropriate for consideration of hierarchical models.

We therefore encourage the following view of capture–recapture models. Data Y are described in terms of their dependence on demographic parameters θ and nuisance parameters η, through distri-
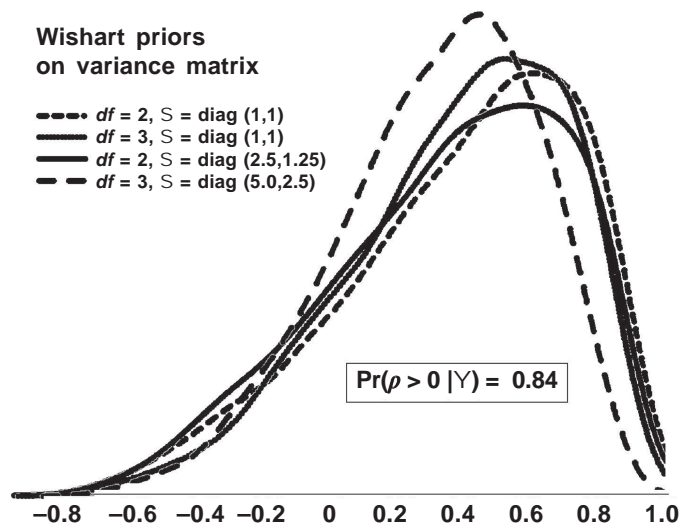

Wishart priors on variance matrix

- *df* = **2**, S = diag (1,1)
- *df* = **3**, S = diag (1,1)
- *df* = **2**, S = diag (2.5,1.25)
- *df* = **3**, S = diag (5.0,2.5)

$Pr(\rho > 0 \,|Y) = 0.84$

−0.8 −0.6 −0.4 −0.2 0 0.2 0.4 0.6 0.8 1.0

Fig. 2. Posterior distributions of ρ.

Fig. 2. Distribución posterior de ρ.

Table 1. Summaries of posterior distributions for *p* in analysis of *Gonodontis* data. First column indicates prior as numbered in text.

Tabla 1. Resúmenes de distribuciones posteriores de p *en el análisis de datos de* Gonodontis. *En la primera columna se indican las distribuciones anteriores según constan numeradas en el texto.*

| | Mean | SD | Quantiles of posterior distribution | | | | | | |
			0.025	0.100	0.250	0.500	0.750	0.900	0.975
1	0.383	0.353	−0.425	−0.131	0.157	0.447	0.664	0.793	0.880
2	0.372	0.323	−0.356	−0.089	0.164	0.420	0.623	0.754	0.850
3	0.345	0.359	−0.450	−0.174	0.108	0.403	0.632	0.772	0.868
4	0.314	0.296	−0.330	−0.098	0.118	0.346	0.54	0.677	0.788

butional assumptions specifying $[Y \mid \theta, \eta]$. The prior distribution $[\theta, \eta]$ is factored as

$$[\theta, \eta] \propto [\eta] \, [\theta \mid \Psi] \; [\Psi]$$

or possibly as

$$[\theta, \eta] \propto [\eta \mid \theta] \, [\theta \mid \Psi] \; [\Psi]$$

if it is thought that the nuisance and demographic parameters are related. Associations among demographic parameters are modeled in the partial specification of the prior, $[\theta \mid \Psi]$.

Given the specification of likelihood and prior, there is no need for asymptotic approximations, no need for selection among multitudinous optimality criteria, no need to "do statistics on statistics".

There remains only a calculation, for the likelihood and prior determine the posterior distribution, upon which all inference is based. Software such as *WinBUGS* (Spiegelhalter et al., 2003) provides an easy entree into Bayesian analysis through implementation of Markov chain Monte Carlo (simulation based) evaluation of the posterior distribution.

References

Bishop, J. A., Cook, L. M. & Muggleton, J., 1978. The response of two species of moths to industrialization in northwest England. II Relative fitness of morphs and populations size. *Phil. Trans. R. Soc. Lond. B.*, 281: 517–540.

Cormack, R. M., 1964. Estimates of survival from the sighting of marked animals. *Biometrika,* 51: 429–438.

Crosbie, S. F., 1979. The mathematical modelling of capture mark recapture experiments on animal populations. Ph. D. Thesis, Univ. Otago, Dunedin, New Zealand.

Crosbie, S. F. & Manly, B. F. J., 1985. Parsimonious modelling of capture mark recapture studies. *Biometrics,* 41: 385–398.

Jolly, G. M., 1965. Explicit estimates from capture recapture data with both death and immigration stochastic model. *Biometrika,* 52: 225–247.

Le Cam, L. M., 1990. Maximum likelihood: an introduction. *International Statistical Review,* 58: 153–171.

Link, W. A., 1999. Modeling pattern in collections of parameters. *Journal of Wildlife Management*, 63: 1017–1027.

Link, W. A. & Barker, R. J. (in press). Modeling association among demographic parameters in analysis of open population capture recapture data, *Biometrics.*

Pradel, R., 1996. Utilization of capture mark recapture for the study of recruitment and population growth rate. *Biometrics,* 52: 703–709.

Schwarz, C. J. & Arnason, A. N., 1996. A general methodology for the analysis of capture recapture experiments in open populations. *Biometrics,* 52: 860–873.

Seber, G. A. F., 1965. A note on the multiple recapture census. *Biometrika,* 52: 249–259.

Spiegelhalter, D., Thomas, A., Best, N. & Lunn, D., 2003. WinBUGS User Manual, version 1.4. http://www.mrc–bsu.cam.ac.uk/bugs

White, G. C. & Burnham, K. P., 1999. Program MARK: Survival estimation from populations of marked animals. *Bird Study*, 46 (Supplement): 120–139.

Appendix / Apéndice

Cell probabilities of the multinomial distribution [$u \mid u.$] are $\zeta_1, \zeta_2, ..., \zeta_t$, defined in terms of $t-1$ estimable parameters $\rho_1, f_2, f_3, ..., f_{t-2}$ and ρ_t. Parameters ρ_1 and ρ_t, like $\lambda_{t-1} = \varphi_{t-1} p_t$ in the CJS model, are confounded combinations of demographic and nuisance parameters. Specifically, these are

$$\rho_1 = (\varphi_1 + f_1) / p_1 \qquad\qquad \text{and} \qquad\qquad \rho_t = f_{t-1} \, p_t.$$

Cell probabilities ζ_i are defined as follows: Let $\delta_2 = \rho_1$ and

$$\delta_i = \rho_1 \prod_{j=2}^{i-1} (\varphi_j + f_j) \qquad \text{for } i = 3, 4, ..., t-1.$$

Let $\kappa_1 \equiv 1$, $\kappa_2 \equiv (\rho_1 - \varphi_1) \, p_2$,

$$\kappa_{i+1} = \frac{\kappa_i}{p_i}(1 - p_i)\varphi_i p_{i+1} + \delta_i f_i p_{i+1} \quad \text{for } i = 2, 3, ..., t-2 \quad \text{and} \quad \kappa_t = \frac{\kappa_{t-1}}{p_{t-1}}(1 - p_{t-1})\lambda_{t-1} + \rho_t \delta_{t-1}$$

Then

$$\xi_i = \kappa_i \bigg/ \sum_{j=1}^{t} \kappa_j$$