2 The capture-recapture models

2.1 A first hypergeometric model

Hypergeometric distribution

$$\mathscr{H}(N, n, p)$$

when the population size, N, rather than p, is unknown

$$\text{Prob}(X = x) = \frac{\binom{pN}{x}\binom{(1-p)N}{n-x}}{\binom{N}{n}}$$

Numerous applications:

- biology and ecology (herds, fish, &tc.),
- sociology and demography (populations at risk, homeless, prostitutes, &tc.),
- official statistics and economics (U.S. Census undercounts),
- fraud detection and document authentification
- software debuggin

Example 26 VCR98054 - Birdwood Mammal Trapping Data, Charlottesville, VA, 1974-1978

Trapping records for 10x10 trapping grids with 7.6m (25 foot) trap spacing. Up to 4 captures of an individual may be recorded on each line of data.

Trapping used modified Fitch live traps with a # 10 tin can as the main chamber. Traps were baited with cracked corn or hen scratch and run for 3 sequential nights in each trapping session.

Variable description

```
GRID Grid
TAG Eartag number
SPECIES Species
FATE1 Fate at capture 1
ROW1 Row of capture
COL1 Grid Column of capture
BMASS    Body Mass
SEX Sex
TESTES Testes Condition
VAGINA Condition of vagina
NIPPLES Size of Nipples
PUBIC Pubic symphysis width
PREG Pregnant?
FATE2 Fate at 2nd capture
ROW2 Row of capture
COL2 Grid Column of Capture
FATE3 Fate at 3rd Capture
ROW3 Row of capture
COL3 Grid Column of Capture
FATE4 Fate at 4th Capture
ROW4 Row of capture
COL4 Grid Column of Capture
WEEK Week number since start of study
```


Darroch Model

$$n_{11} \sim \mathscr{H}(N, n_2, n_1/N)$$

Classical (MLE) estimator of N

$$\hat{N} = \frac{n_1}{(n_{11}/n_2)},$$

Important drawback:

It cannot be used when $n_{11} = 0$

Example 27 Deers

Herd of deer on an island of Newfoundland (Canada) w/o any predator.

Culling necessary for ecological equilibrium.

Annual census too time-consuming, but birth and death patterns for the deer imply that the number of deer varies between 36 and 50. Prior:

$$N \sim \mathscr{U}(\{36, \dots, 50\})$$

Posterior distribution

$$\pi(N = n|n_{11}) = \frac{\binom{n_1}{n_{11}}\binom{n_2}{n_2 - n_{11}} \Big/ \binom{n}{n_2} \pi(N = n)}{\sum_{k=36}^{50} \binom{n_1}{n_{11}}\binom{n_2}{n_2 - n_{11}} \Big/ \binom{k}{n_2} \pi(N = k)},$$

Table 4: Posterior distribution of the deer population size, $\pi(N|n_{11})$.

$N \backslash n_{11}$	0	1	2	3	4	5
36	0.058	0.072	0.089	0.106	0.125	0.144
37	0.059	0.072	0.085	0.098	0.111	0.124
38	0.061	0.071	0.081	0.090	0.100	0.108
39	0.062	0.070	0.077	0.084	0.089	0.094
40	0.063	0.069	0.074	0.078	0.081	0.082
41	0.065	0.068	0.071	0.072	0.073	0.072
42	0.066	0.068	0.067	0.067	0.066	0.064
43	0.067	0.067	0.065	0.063	0.060	0.056
44	0.068	0.066	0.062	0.059	0.054	0.050
45	0.069	0.065	0.060	0.055	0.050	0.044
46	0.070	0.064	0.058	0.051	0.045	0.040
47	0.071	0.063	0.056	0.048	0.041	0.035
48	0.072	0.063	0.054	0.045	0.038	0.032
49	0.073	0.062	0.052	0.043	0.035	0.028
50	0.074	0.061	0.050	0.040	0.032	0.026

Table 5: Posterior mean of the deer population size, N.

n_{11}	0	1	2	3	4	5
$\mathbb{E}(N|n_{11})$	43.32	42.77	42.23	41.71	41.23	40.78

Different loss function

$$\mathrm{L}(N, \delta) = \begin{cases} 10(\delta - N) & \text{if } \delta > N, \\ N - \delta & \text{otherwise,} \end{cases}$$

in order to avoid overestimation

Bayes estimator is $(1/11)$-quantile of $\pi(N|n_{11})$,

Table 6: Estimated deer population

n_{11}	0	1	2	3	4	5
$\delta^\pi(n_{11})$	37	37	37	36	36	36

Darroch model (2)

Unknown capture probability p

$$\mathsf{L}(N, p|\mathscr{D}) = \prod_t \prod_i p^{\delta_{it}} (1 - p)^{1 - \delta_{it}}$$

with δ_{it} capture indicator

Equivalent to

$$\mathsf{L}(N,p|\mathscr{D}) = \binom{N}{n_1 \ldots n_T} p^{n_1+\ldots+Tn_T}(1-p)^{TN-n_1-\ldots-Tn_T}$$
$$\propto \frac{N!}{(N-n^+)!} p^{n^c}(1-p)^{TN-n^c}$$

where n^+ number of captured individuals and n^c number of captures

Also equivalent to cascade sampling:

$$n_1 \sim \mathscr{B}(N,p), \quad n_2 \sim \mathscr{B}(n_1,p), \ldots$$

For a prior

$$\pi(N,p) = 1\Big/\sqrt{p(1-p)}\,N$$

posterior

$$\pi(N,p|\mathscr{D}) \propto \frac{(N-1)!}{(N-n^+)!} p^{n^c-1/2}(1-p)^{N-n^c-1/2}$$

with conditionnal distributions

$$\pi(N|p,\mathscr{D}) \propto \frac{(N-1)!}{(N-n_1)!}(1-p)^N \mathbb{I}_{N\geq n^+} \,,$$
$$\pi(p|N,\mathscr{D}) \propto p^{n^c-1/2}(1-p)^{TN-n^c-1/2} \,,$$

and

$$p|N,\mathscr{D} \sim \mathscr{B}e(n^c+1/2, TN-n^c+1/2) \,,$$
$$N \sim \pi(N|\mathscr{D}) \propto \frac{(N-1)!}{(N-n^+)!}\frac{\Gamma(TN-n^c+1/2)}{(TN+1)!}\mathbb{I}_{N\geq n^+}$$

[Computable!!!]

2.2 A more advanced sampling model

Heterogeneous *capture–recapture model* :

Animals captured at time i with both probability p_i and size N of the population unknown.

Example 28 **Northern Pintail ducks**

Dataset

$$(n_1,\ldots,n_{11}) = (32,20,8,5,1,2,0,2,1,1,0)$$

Number of recoveries over the years 1957–1968 of $N = 1612$ Northern Pintail ducks banded in 1956

Corresponding likelihood

$$L(p_1,\ldots,p_I|N,n_1,\ldots,n_I) = \frac{N!}{(N-r)!}\prod_{i=1}^{I} p_i^{n_i}(1-p_i)^{N-n_i},$$

where I number of captures, n_i number of captured animals during the ith capture, and r is the total number of different captured animals.

Prior selection

If

$$N \sim \mathcal{P}(\lambda)$$

and

$$\alpha_i = \log\left(\frac{p_i}{1-p_i}\right) \sim \mathcal{N}(\mu_i,\sigma^2),$$

[Normal logistic]

Posterior distribution

then

$$\pi(\alpha, N|, n_1, \ldots, n_I) \quad \propto \quad \frac{N!}{(N-r)!}\frac{\lambda^N}{N!}\prod_{i=1}^{I}(1+e^{\alpha_i})^{-N}$$
$$\prod_{i=1}^{I}\exp\left\{\alpha_i n_i - \frac{1}{2\sigma^2}(\alpha_i-\mu_i)^2\right\}$$

Just too hard to work with!!!

2.2.1 Accept-Reject Methods

- Many distributions from which difficult, or even impossible, to **directly** simulate.
- Another class of methods that only require us to know the functional form of the density f of interest **only** up to a multiplicative constant.
- The key to this method is to use a simpler (simulation-wise) density g, the *instrumental density*, from which the simulation from the *target density* f is actually done.

Fundamental theorem of simulation

Simulating

$$X \sim f(x)$$

equivalent to simulating

$$(X, U) \sim \mathcal{U}\{(x, u) : 0 < u < f(x)\}$$

Accept-Reject method

Given a density of interest f, find a density g and a constant M such that

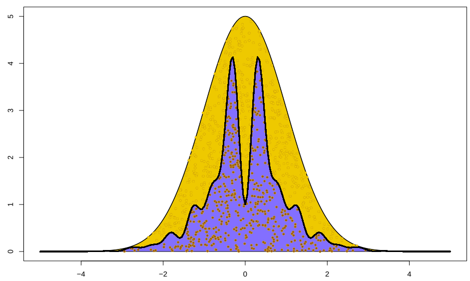
$$f(x) \le Mg(x)$$

on the support of f.

1. Generate $X \sim g$, $U \sim \mathcal{U}_{[0,1]}$;
2. Accept $Y = X$ if $U \le f(X)/Mg(X)$;
3. Return to 1. otherwise.

Validation of the Accept-Reject method

This algorithm produces a variable Y distributed according to f



Uniform repartition under the graph of f of accepted points

Two interesting properties:

- First, it provides a generic method to simulate from any density f that is known *up to a multiplicative factor*

  Property particularly important in Bayesian calculations: there, the posterior distribution

  $$\pi(\theta|x) \propto \pi(\theta)\ f(x|\theta)\ .$$

  is specified up to a normalizing constant

- Second, the probability of acceptance in the algorithm is $1/M$, e.g., expected number of trials until a variable is accepted is M

Log-concave densities

Densities f whose logarithm is concave, for instance Bayesian posterior distributions such that

$$\log\ \pi(\theta|x) = \log\ \pi(\theta) + \log\ f(x|\theta) + c$$

concave

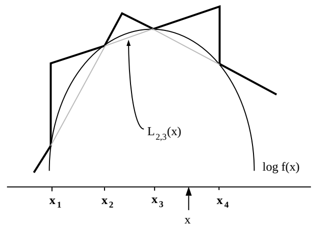
Take

$$\mathcal{S}_n = \{x_i, i = 0, 1, \ldots, n+1\} \subset \text{supp}(f)$$

such that $h(x_i) = \log f(x_i)$ known up to the same constant.

By concavity of h, line $L_{i,i+1}$ through $(x_i, h(x_i))$ and $(x_{i+1}, h(x_{i+1}))$

- below h in $[x_i, x_{i+1}]$ and
- above this graph outside this interval



For $x \in [x_i, x_{i+1}]$, if

$$\overline{h}_n(x) = \min\{L_{i-1,i}(x), L_{i+1,i+2}(x)\} \quad \text{and} \quad \underline{h}_n(x) = L_{i,i+1}(x)\ ,$$

the envelopes are

$$\underline{h}_n(x) \le h(x) \le \overline{h}_n(x)$$

uniformly on the support of f, with

$$\underline{h}_n(x) = -\infty \quad \text{and} \quad \overline{h}_n(x) = \min(L_{0,1}(x), L_{n,n+1}(x))$$

on $[x_0, x_{n+1}]^c$. Therefore, if

$$\underline{f}_n(x) = \exp \underline{h}_n(x) \text{ and } \overline{f}_n(x) = \exp \overline{h}_n(x)$$

then

$$\underline{f}_n(x) \le f(x) \le \overline{f}_n(x) = \varpi_n\ g_n(x)\ ,$$

where ϖ_n normalizing constant of f_n

Algorithm 29 –ARS Algorithm–

0. Initialize n and $\mathcal{S}_n$.
1. Generate $X \sim g_n(x)$, $U \sim \mathcal{U}_{[0,1]}$.
2. If $U \le \underline{f}_n(X)/\varpi_n\ g_n(X)$, accept X; otherwise, if $U \le f(X)/\varpi_n\ g_n(X)$, accept X and update $\mathcal{S}_n$ to $\mathcal{S}_{n+1} = \mathcal{S}_n \cup \{X\}$.

Example 30 **Northern Pintail ducks**

For the posterior distribution

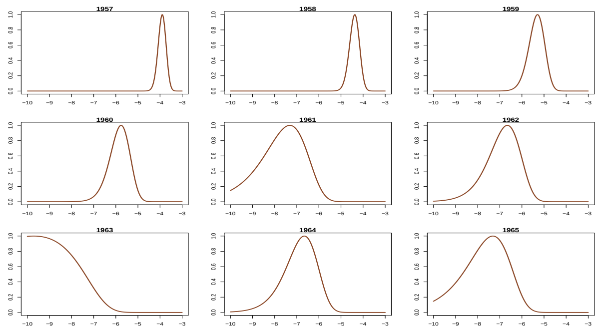
$$\pi(\alpha_i|N,n_1,\ldots,n_I) \;\propto\; \exp\left\{\alpha_i n_i - \frac{1}{2\sigma^2}(\alpha_i-\mu_i)^2\right\}\Big/(1+e^{\alpha_i})^N\,,$$

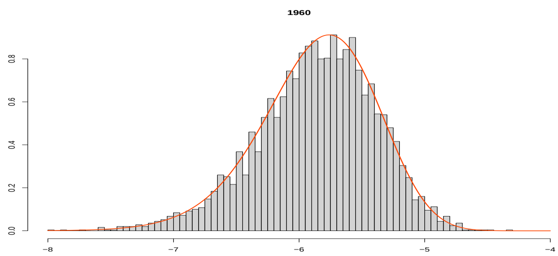
the ARS algorithm can be implemented since

$$\alpha_i n_i - \frac{1}{2\sigma^2}\,(\alpha_i-\mu_i)^2 - N\,\log(1+e^{\alpha_i})$$

is concave in α_i.

Posterior distributions of capture log-odds ratios for the years 1957–1965.





True distribution versus histogram of simulated sample

2.2.2 Monte Carlo methods

Approximation of the integral

$$\mathfrak{I} = \int_\Theta g(\theta) f(x|\theta)\pi(\theta)\, d\theta,$$

should take advantage of the fact that $f(x|\theta)\pi(\theta)$ is proportional to a density.

If the θ_i's are generated from $\pi(\theta)$, the average

$$\frac{1}{m}\sum_{i=1}^m g(\theta_i) f(x|\theta_i)$$

converges (almost surely) to $\mathfrak{I}$

Confidence regions can be derived from a normal approximation and the magnitude of the error remains of order

$$1/\sqrt{m}\,,$$

whatever the dimension of the problem.

Importance function

No need to simulate from $\pi(\cdot|x)$ or π: if h is a probability density,

$$\int_\Theta g(\theta) f(x|\theta)\pi(\theta)\, d\theta = \int \frac{g(\theta) f(x|\theta)\pi(\theta)}{h(\theta)} h(\theta)\, d\theta.$$

[Importance function]

An approximation to $\mathbb{E}^\pi[g(\theta)|x]$ is given by

$$\frac{\sum_{i=1}^m g(\theta_i)\omega(\theta_i)}{\sum_{i=1}^m \omega(\theta_i)} \quad \text{with} \quad \omega(\theta_i) = \frac{f(x|\theta_i)\pi(\theta_i)}{h(\theta_i)}$$

if

$$\text{supp}(h) \subset \text{supp}(f(x|\cdot)\pi)$$

Requirements

- Simulation from h must be easy
- $h(\theta)$ must be close enough to $g(\theta)\pi(\theta|x)$
- the variance of the importance sampling estimator must be finite

The importance function may be π

Example 31 Consider

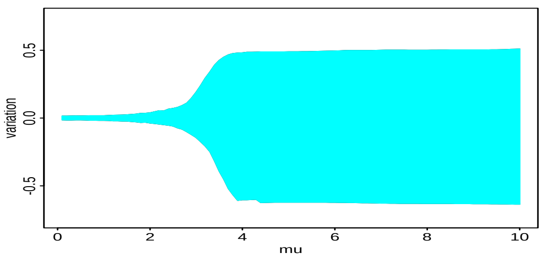
$$x_1, \ldots, x_n \sim \mathcal{C}(\theta, 1)$$

and $\theta \sim \mathcal{N}(\mu, \sigma^2)$, with known hyperparameters μ and σ^2.

Since $\pi(\theta)$ is the normal distribution $\mathcal{N}(\mu, \sigma^2)$, it is possible to simulate a normal sample $\theta_1, \ldots, \theta_M$ and to approximate the Bayes estimator by

$$\hat{\delta}^\pi(x_1, \ldots, x_n) = \frac{\sum_{t=1}^M \theta_t \prod_{i=1}^n [1 + (x_i - \theta_t)^2]^{-1}}{\sum_{t=1}^M \prod_{i=1}^n [1 + (x_i - \theta_t)^2]^{-1}}.$$

May be poor when the x_i's are all far from μ



90% range of variation for $n = 10$ observations from $\mathcal{C}(0, 1)$ distribution and $M = 1000$ simulations of θ from a $\mathcal{N}(\mu, 1)$ distribution.

Defensive sampling:

$$h(\theta) = \rho\pi(\theta) + (1 - \rho)\pi(\theta|x) \qquad \rho \ll 1$$

[Newton & Raftery, 1994]

Case of the Bayes factor

Models $\mathcal{M}_1$ vs. $\mathcal{M}_2$ compared via

$$B_{12} = \frac{Pr(\mathcal{M}_1|x)}{Pr(\mathcal{M}_2|x)} \Big/ \frac{Pr(\mathcal{M}_1)}{Pr(\mathcal{M}_2)} = \frac{\int f_1(x|\theta_1)\pi_1(\theta_1)d\theta_1}{\int f_2(x|\theta_2)\pi_2(\theta_2)d\theta_2}$$

[Good, 1958 & Jeffreys, 1961]

Solutions

- **Bridge sampling:**

  If

$$\pi_1(\theta_1|x) \propto \tilde{\pi}_1(\theta_1|x)$$
$$\pi_2(\theta_2|x) \propto \tilde{\pi}_2(\theta_2|x)$$

  then

$$B_{12} \approx \frac{1}{n}\sum_{i=1}^n \frac{\tilde{\pi}_1(\theta_i|x)}{\tilde{\pi}_2(\theta_i|x)} \qquad \theta_i \sim \pi_2(\theta|x)$$

[Chen, Shao & Ibrahim, 2000]

- **Umbrella sampling:**

$$\pi_1(\theta) = \pi(\theta|\lambda_1) = \tilde{\pi}_1(\theta)/c(\lambda_1) \qquad \pi_2(\theta) = \pi_1(\theta|\lambda_2) = \tilde{\pi}_2(\theta)/c(\lambda_2)$$

Then

$$\forall\ \pi(\lambda) \text{ on } [\lambda_1, \lambda_2], \qquad \log(c(\lambda_2)/c(\lambda_1)) = \mathbb{E}\left[\frac{\frac{d}{d\lambda}\log\tilde{\pi}(d\theta)}{\pi(\lambda)}\right]$$

and

$$\log(B_{12}) \approx \frac{1}{n}\sum_{i=1}^{n}\frac{\frac{d}{d\lambda}\log\tilde{\pi}(\theta_i|\lambda_i)}{\pi(\lambda_i)}$$

2.3 Markov chain Monte Carlo methods

Idea **Given a density distribution $\pi(\cdot|x)$, produce a Markov chain $(\theta^{(t)})_t$ with stationary distribution $\pi(\cdot|x)$**

Warranty:

if the Markov chains produced by MCMC algorithms are irreducible, then these chains are positive recurrent with stationary distribution $\pi(\theta|x)$ and ergodic.

Translation:
For k large enough, $\theta^{(k)}$ is approximately distributed from $\pi(\theta|x)$, no matter what the starting value $\theta^{(0)}$ is.

Practical use

- Produce an i.i.d. sample $\theta_1, \ldots, \theta_m$ from $\pi(\theta|x)$, taking the current $\theta^{(k)}$ as the new starting value
- Approximate $\mathbb{E}^\pi[g(\theta)|x]$ as

$$\frac{1}{K}\sum_{k=1}^{K} g(\theta^{(k)})$$

[Ergodic Theorem]

- Achieve quasi-independence by batch sampling
- Construct approximate posterior confidence regions

$$C_x^\pi \simeq [\theta^{(\alpha T/2)}, \theta^{(T-\alpha T/2)}]$$

2.3.1 The Gibbs sampler

Takes advantage of *hierarchical structures*: if

$$\pi(\theta|x) = \int \pi_1(\theta|x,\lambda)\pi_2(\lambda|x)\,d\lambda\,,$$

simulate from the joint distribution

$$\pi_1(\theta|x,\lambda)\ \pi_2(\lambda|x)$$

Example 32 Consider $(\theta,\lambda) \in \mathbb{N}\times[0,1]$ and

$$\pi(\theta,\lambda|x) \propto \binom{n}{\theta}\lambda^{\theta+\alpha-1}(1-\lambda)^{n-\theta+\beta-1}$$

Hierarchical structure:

$$\theta|x,\lambda \sim \mathcal{B}(n,\lambda), \qquad \lambda|x \sim \mathcal{B}e(\alpha,\beta)$$

Then

$$\pi(\theta|x) = \binom{n}{\theta}\frac{B(\alpha+\theta,\beta+n-\theta)}{B(\alpha,\beta)}$$

[beta-binomial distribution]

Difficult to work with this marginal : For instance, computation of $\mathbb{E}[\theta/(\theta+1)|x]$?

More advantageous to simulate

$$\lambda^{(i)} \sim \mathcal{B}e(\alpha,\beta) \text{ and } \theta^{(i)} \sim \mathcal{B}(n,\lambda^{(i)})$$

Then approximate $\mathbb{E}[\theta/(\theta+1)|x]$ as

$$\frac{1}{m}\sum_{i=1}^{m}\frac{\theta^{(i)}}{\theta^{(i)}+1}$$

Conditionals

Usually $\pi_2(\lambda|x)$ not available/simulable

More often, both *conditional posterior distributions*,

$$\pi_1(\theta|x,\lambda) \text{ and } \pi_2(\lambda|x,\theta)$$

can be simulated.

Example 33 For the capture-recapture model, the two conditional posterior distributions are $(1 \le i \le I)$

$$\begin{aligned} p_i|x,N &\sim \mathcal{B}e(\alpha+x_i,\beta+N-x_i) \\ N-x_+|x,p &\sim \mathcal{N}eg(x_+,\varrho), \end{aligned}$$

with

$$\varrho = 1-\prod_{i=1}^{I}(1-p_i).$$

Data augmentation algorithm

Initialization: Start with an arbitrary value $\lambda^{(0)}$

Iteration t: Given $\lambda^{(t-1)}$, generate

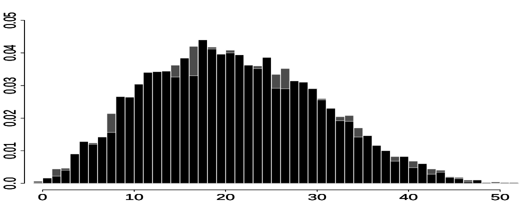
a. $\theta^{(t)}$ according to $\pi_1(\theta|x,\lambda^{(t-1)})$

b. $\lambda^{(t)}$ according to $\pi_2(\lambda|x,\theta^{(t)})$

$\pi(\theta,\lambda|x)$ is a stationary distribution for this transition

Example 34 (Example 32 continued) The conditional distributions are

$$\theta|x,\lambda \sim \mathcal{B}(n,\lambda), \qquad \lambda|x,\theta \sim \mathcal{B}e(\alpha+\theta,\beta+n-\theta)$$



Histograms for samples of size 5000 from the beta-binomial with $n=54$, $\alpha=3.4$, and $\beta=5.2$

Rao–Blackwellization

Conditional structure of the sampling algorithm and the dual sample,

$$\lambda^{(1)},\ldots,\lambda^{(m)},$$

should be exploited.

$\mathbb{E}^\pi[g(\theta)|x]$ approximated as

$$\delta_2 = \frac{1}{m}\sum_{i=1}^{m}\mathbb{E}^\pi[g(\theta)|x,\lambda^{(m)}],$$

instead of

$$\delta_1 = \frac{1}{m}\sum_{i=1}^{m} g(\theta^{(i)}).$$

Approximation of $\pi(\theta|x)$ by

$$\frac{1}{m}\sum_{i=1}^{m} \pi(\theta|x, \lambda_i)$$

The general Gibbs sampler

Consider several groups of parameters, $\theta, \lambda_1, \ldots, \lambda_p$, such that

$$\pi(\theta|x) = \int \ldots \int \pi(\theta, \lambda_1, \ldots, \lambda_p|x)\, d\lambda_1 \cdots d\lambda_p$$

or simply divide θ in

$$(\theta_1, \ldots, \theta_p)$$

Example 35 Consider a multinomial model,

$$y \sim \mathcal{M}_5\left(n; a_1\mu + b_1, a_2\mu + b_2, a_3\eta + b_3, a_4\eta + b_4, c(1-\mu-\eta)\right),$$

parametrized by μ and η, where

$$0 \le a_1 + a_2 = a_3 + a_4 = 1 - \sum_{i=1}^{4} b_i = c \le 1$$

and $c, a_i, b_i \ge 0$ are known.

This model stems from sampling according to

$$x \sim \mathcal{M}_9(n; a_1\mu, b_1, a_2\mu, b_2, a_3\eta, b_3, a_4\eta, b_4, c(1-\mu-\eta)),$$

and aggregating some coordinates:

$$y_1 = x_1 + x_2, \quad y_2 = x_3 + x_4, \quad y_3 = x_5 + x_6, \quad y_4 = x_7 + x_8,\ y_5 = x_9.$$

For the prior

$$\pi(\mu, \eta) \propto \mu^{\alpha_1 - 1}\eta^{\alpha_2 - 1}(1-\eta-\mu)^{\alpha_3 - 1},$$

the posterior distribution of (μ, η) cannot be derived explicitly.

Introduce $z = (x_1, x_3, x_5, x_7)$, which is not observed and

$$\begin{aligned} \pi(\eta, \mu|y, z) &= \pi(\eta, \mu|x) \\ &\propto \mu^{z_1}\mu^{z_2}\eta^{z_3}\eta^{z_4}(1-\eta-\mu)^{y_5+\alpha_3-1}\mu^{\alpha_1-1}\eta^{\alpha_2-1}, \end{aligned}$$

where we denote the coordinates of z as (z_1, z_2, z_3, z_4). Therefore,

$$\mu, \eta|y, z \sim \mathcal{D}(z_1 + z_2 + \alpha_1, z_3 + z_4 + \alpha_2, y_5 + \alpha_3).$$

Moreover,

$$\begin{aligned} z_i|y, \mu, \eta &\sim \mathcal{B}\left(y_i, \frac{a_i\mu}{a_i\mu + b_i}\right) \qquad (i = 1, 2), \\ z_i|y, \mu, \eta &\sim \mathcal{B}\left(y_i, \frac{a_i\eta}{a_i\eta + b_i}\right) \qquad (i = 3, 4). \end{aligned}$$

The Gibbs sampler

For a joint distribution $\pi(\theta)$ with full conditionals $\pi_1, \ldots, \pi_p$,

Given $(\theta_1^{(t)}, \ldots, \theta_p^{(t)})$, simulate

1. $\theta_1^{(t+1)} \sim \pi_1(\theta_1|\theta_2^{(t)}, \ldots, \theta_p^{(t)})$,
2. $\theta_2^{(t+1)} \sim \pi_2(\theta_2|\theta_1^{(t+1)}, \theta_3^{(t)}, \ldots, \theta_p^{(t)})$,

$\vdots$

p. $\theta_p^{(t+1)} \sim \pi_p(\theta_p|\theta_1^{(t+1)}, \ldots, \theta_{p-1}^{(t+1)})$.

Example 36 **Open population model**

Probability q to leave the population each time

$$\mathcal{L}(N, p|\mathcal{D}^*) = \prod_t \prod_i q_{\epsilon_{i(t-1)}}^{\epsilon_{it}} (1 - q_{\epsilon_{i(t-1)}})^{1-\epsilon_{it}}$$
$$p^{(1-\epsilon_{it})\delta_{it}} (1-p)^{(1-\epsilon_{it})(1-\delta_{it})}$$

where $q_0 = q$, $q_1 = 1$, and ϵ_{it} exit indicator.

Substitution model

$$n_1 \sim \mathcal{B}(N, p), \quad n_2 \sim \mathcal{B}(n_1, pq), \quad n_3 \sim \mathcal{B}(n_2, pq), \ldots$$

Associated conditionals

$$\pi(p|N, q, \mathcal{D}) \propto p^{n_+ - 1/2}(1-p)^{N-n_1-1/2}(1-pq)^{n_1-n_k}$$
$$\pi(q|N, p, \mathcal{D}) \propto q^{n_+ - n_1 + \alpha - 1}(1-q)^{\beta-1}(1-pq)^{n_1-n_k}$$
$$\pi(N|p, q, \mathcal{D}) \propto \frac{(N-1)!}{(N-n_1)!}(1-p)^N \mathbb{I}_{N \geq n_1}$$

and

$$p|N, q, \mathcal{D} \sim \mathcal{B}e(n_+ + 1/2, N - n_1 + 1/2 + q(n_1 - n_k))$$
$$q|N, p, \mathcal{D} \sim \mathcal{B}e(n_+ - n_1 + \alpha, \beta + p(n_1 - n_k))$$
$$N - n_1|p, q, \mathcal{D} \sim \mathcal{P}oi((1-p)n_1/p)$$

for substitution model

2.3.2 The impact of MCMC on Bayesian Statistics

- Radical modification of the way people work with models and prior assumptions
- Allows for much more complex structures:
  - use of graphical models
  - exploration of latent variable models
- Removes the need for analytical processing
- Boosted hierarchical modeling
- Enables *(truly)* Bayesian model choice

2.4 An even more advanced capture-recapture model

2.4.1 Arnason–Schwarz model

Estimate movement and survival probabilities for individuals

Example 37 Study a zone K divided in $k = 3$ strata a, b, c

Four possible capture-recapture histories:

a	c	·	·	b	a	b	c
a	a	c	·	·	a	a	b
·	c	a	c	b	a	a	·
·	a	a	c	a	a	·	·

where " · " denotes a failure of capture

Missing data structure

- $z_{(i,t)} = r$: the animal i is (alive) in stratum r at time t;
- $z_{(i,t)} = \dagger$: the animal i is dead at time t.
- $\mathbf{z}_i = (z_{(i,t)}, t = 1, .., \tau)$ **migration** process related to i.
- $x_{(i,t)} = 0$: failure of capture of i at time t (the location $z_{(i,t)}$ is missing)
- $\mathbf{z}_i = (x_{(i,t)}, t = 1, .., \tau)$ **capture** process related to i.
- $\mathbf{y}_i$ **capture-recapture history** of animal i.

Example 38

$$\mathbf{y}_i = 1\,2\,\cdot\,3\,1\,1\,\cdot\,\cdot\,\cdot$$

For this capture-recapture history we have

$$\mathbf{x}_i = 1\,1\,0\,1\,1\,1\,0\,0\,0$$

A possible $\mathbf{z}_i$ is

$$\mathbf{z}_i = 1\,2\,1\,3\,1\,1\,\dagger\,\dagger$$

Parameters of the Arnason-Schwarz model

Capture probabilities

$$p_t(r) = \text{Pr}\left(x_{(i,t)} = 1 | z_{(i,t)} = r\right)$$

Transition probabilities

$$q_t(r,s) = \text{Pr}\left(z_{(i,t+1)} = s | z_{(i,t)} = r\right) \quad r \in K, s \in K \cup \{\dagger\}$$

Survival and movement probabilities

$$q_t(r,s) = \phi_t(r) \times \psi_t(r,s) \quad r \in K, s \in K$$

$\phi_t(r) = 1 - q_t(r,\dagger)$ **survival** probability. $\psi_t(r,s)$ inter-strata **movement** probability.

2.4.2 Prior modelling

Conjugate priors

$$p_t(r) \sim \mathscr{B}e(a_t(r), b_t(r)) \qquad \phi_t(r) \sim \mathscr{B}e(\alpha_t(r), \beta_t(r))$$

and

$$\psi_t(r) \sim \mathscr{D}ir(\gamma_t(r))$$

where $\psi_t(r) = (\psi_t(r,s); s = 1, \ldots, k)$ with

$$\sum_{s=1}^{k} \psi_t(r,s) = 1$$

and $\gamma_t(r) = (\gamma_t(r,s); s = 1, \ldots, k)$.

Example 39 **Capture-recapture experiment on migrations between zones**

Prior information on capture and survival probabilities, p_t and q_{it}

	Time	2	3	4	5	6
p_t	Mean	0.3	0.4	0.5	0.2	0.2
	95% cred. int.	[0.1,0.5]	[0.2,0.6]	[0.3,0.7]	[0.05,0.4]	[0.05,0.4]

	Site	A		B	
	Time	t=1,3,5	t=2,4	t=1,3,5	t=2,4
q_{it}	Mean	0.7	0.65	0.7	0.7
	95% cred. int.	[0.4,0.95]	[0.35,0.9]	[0.4,0.95]	[0.4,0.95]

Corresponding prior modeling

Time	2	3	4	5	6
Dist.	$\mathcal{B}e(6,14)$	$\mathcal{B}e(8,12)$	$\mathcal{B}e(12,12)$	$\mathcal{B}e(3.5,14)$	$\mathcal{B}e(3.5,14)$

Site	A		B	
Time	t=1,3,5	t=2,4	t=1,3,5	t=2,4
Dist.	$\mathcal{B}e(6.0,2.5)$	$\mathcal{B}e(6.5,3.5)$	$\mathcal{B}e(6.0,2.5)$	$\mathcal{B}e(6.0,2.5)$

2.4.3 Gibbs sampling

Advantage of using the missing data structure

$$\pi(|\mathbf{y},) \propto L(|\mathbf{y},) \times \pi()$$

simple and easily simulated, thanks to conjugacy

At iteration t

1 **Parameter simulation**

simulate $\theta^{(l)} \sim \pi(\theta|^{(l-1)}, \mathbf{y})$ as

$$p_t^{(l)}(r)|(\mathbf{y},^{(l-1)}) \sim \mathscr{B}e(a_t(r) + u_t(r), b_t(r) + v_t^{(l)}(r))$$

$$\phi_t^{(l)}(r)|(\mathbf{y},^{(l-1)}) \sim \mathscr{B}e(\alpha_t(r) + \sum_j w_t^{(l)}(r,j), \beta_t(r) + w_t^{(l)}(r,\dagger))$$

$$\psi_t^{(l)}(r)|(\mathbf{y},^{(l-1)}) \sim \mathscr{D}ir(\gamma_t(r,s) + w_t^{(l)}(r,s); s \in K)$$

where

$$w_t^{(l)}(r,s) = \sum_{t=1}^{\tau-1} \mathbb{I}_{(z_{(i,t)}=r, z_{(i,t+1)}=s)}$$

$$u_t^{(l)}(r) = \sum_{t=1}^{\tau} \mathbb{I}_{(x_{(i,t)}=1, z_{(i,t)}=r)}$$

$$v_t^{(l)}(r) = \sum_{t=1}^{\tau} \mathbb{I}_{(x_{(i,t)}=0, z_{(i,t)}=r)}$$

and where $w_t^{(l)}(r,.) = w_t^{(l)}(r,s)$

2 **Missing data simulation**

generate the $z_{(i,t)}$'s as

$$p(z_{(i,t)}|x_{(i,t-1)}, z_{(i,t-1)}, z_{(i,t+1)},^{(l)})$$

Example 40 Take $K = \{1,2\}, m = 8$ and, for $\mathbf{y}$

1	1	·	·	1	·	·	·
1	·	1	·	1	·	2	1
2	1	·	1	2	·	·	1
1	·	·	1	2	1	1	2

For instance, at step $(l-1)$, completed data $(\mathbf{y},^{(l-1)})$

1	1	1	2	1	1	2	†
1	1	1	2	1	1	1	2
2	1	2	1	2	1	1	1
1	2	1	1	2	1	1	2

then simulation parameter phase as follows:

$$p_4^{(l)}(1)|(\mathbf{y},^{l-1}) \sim \mathscr{B}e(1+2, 1+2)$$

$$\phi_4^{(l)}(2)|(\mathbf{y},^{l-1}) \sim \mathscr{B}e(1+4, 1+0)$$

$$\psi_4^{(l)}(1,2)|(\mathbf{y},^{(l-1)}) \sim \mathscr{B}e(1+2, 1+1)$$