Capture-Recapture in a salad bowl (and extensions) Frequentist vs Bayesian Estimations

Éric Parent & Etienne Rivot
AgroParisTech, UMR518, Dpt of Maths, Paris
AgroCampus Ouest, UMR0985, Dpt of Ecology, Rennes
Eric.Parent@agroparistech.fr
Etienne.Rivot@agrocampus-ouest.fr

Sunday the 29th of June 2014

Agenda

Conceptual basis

- Examples and a toy (but realistic) example
- Write model, data and assess phenomenological knowledge
- Frequentist vs Bayesian interpretations

Side effect review of some mathematical tricks in estimation and conjugate families

- The exponential family toolbox
- Where does the prior comes from?
- Bayesian computation algorithms

Context

- Census = exhaustive counting of individuals belonging to a (statistical) population.
- When impossible $\Longrightarrow$ Capture - mark - recapture models.

Examples

- Some forest manager would like to know the number of oaks in the forest he is in charge of.
- A fishing society is interested in the number of trouts leaving in some portion of the river.
- The city of Glasgow wishes to evaluate the number of prostitutes in order to increase knowledge about the risks of infectuous disease propagations.

Unknowns and observables

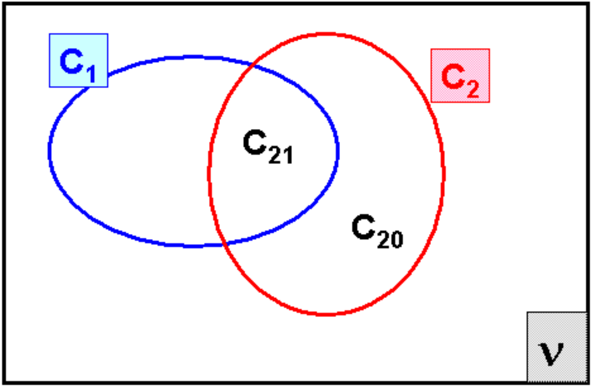
Use Greek letters

- ν population size
- π_1 capture prob experiment 1
- π_2 capture prob experiment 2

And Latin ones....

- C_1first catch (marked)
- C_{20} unmarked part of the second catch
- C_{21} already marked part during first experiment of the second catch
- $C_2 = C_{20} + C_{21}$

A little drawing may be more informative than a long speech



Pick your own

In a salad bowl...

- The tablespoon is your "fishing" device.
- A salad bowl (30 to 40 cm diameter) or a metal cake shape ($30 \times 10 \times 10$ cm3) as the lake.
- One kg of rice to simulate the lake waters.
- A bag of dried beans, e. g. 500g of white or red beans; here are the fish!
- A pack of adhesive labels from various colors are used to mark caught fish.
- Objective: make your own estimation of the number of individuals and give a confidence interval
- NB: in practice, sweep up at least six times to gather enough fish . One third pass might provide additional information.

Let's discuss

- Hypotheses
- Likelihood & independence
- What is a model?
- What is an estimate?

Models

Balance budget under a global point of view

	Both times	1^{st} catch only	2^{nd} catch only	never
Prob.	$\pi_1\pi_2$	$\pi_1(1-\pi_2)$	$(1-\pi_1)\pi_2$	$(1-\pi_1)(1-\pi_2)$
Number	C_{21}	$C_1 - C_{21}$	C_{20}	$\nu - (C_1 - C_{21} + C$
Coding?	11	10	01	00

Constructive step by step point of view

$$\begin{aligned} C_1 &\sim rbinom(1, \nu, \pi_1) \\ C_{20}|C_1 &\sim rbinom(1, \nu - C_1, \pi_2) \\ C_{21}|C_1 &\sim rbinom(1, C_1, \pi_2) \end{aligned}$$

According to global binomial events

$$
\begin{aligned}
[C_1, C_{20}, C_{21}] &= [C_{20}, C_{21} | C_1][C_1] \\
&= [C_{20} | C_1][C_{21} | C_1][C_1] \\
&= \frac{\pi_1^{C_1}(1-\pi_1)^{\nu - C_1}\nu!}{(\nu - C_1)!C_1!} \\
&\times \frac{\pi_2^{C_{20}}(1-\pi_2)^{\nu - C_1 - C_{20}}(\nu - C_1)!}{(\nu - C_1 - C_{20})!C_{20}!} \\
&\times \frac{\pi_2^{C_{21}}(1-\pi_2)^{C_1 - C_{21}}C_1!}{(C_1 - C_{21})!C_{21}!}
\end{aligned}
$$

$$
\begin{aligned}
[C_1, C_{20}, C_{21}] &= \frac{\pi_1^{C_1}(1-\pi_1)^{\nu - C_1}\nu!}{(\nu - C_1)!C_1!} \times \frac{\pi_2^{C_{20}}(1-\pi_2)^{\nu - C_1 - C_{20}}(\nu - C_1)!}{(\nu - C_1 - C_{20})!C_{20}!} \\
&\times \frac{\pi_2^{C_{21}}(1-\pi_2)^{C_1 - C_{21}}C_1!}{(C_1 - C_{21})!C_{21}!} \\
&= \frac{\nu!}{C_{21}!(C_1 - C_{21})!C_{20}!(\nu - C_1 - C_{20})!} \\
&\times \pi_1^{C_1 - C_{21} + C_{21}}(1-\pi_1)^{\nu - C_1 + C_{21} - C_2 + C_{20}}\pi_2^{C_{20}+C_{21}}(1-\pi_2)^{C_1 - C_1 +} \\
&\propto (\pi_1\pi_2)^{C_{21}}\left(\pi_1(1-\pi_2)\right)^{C_1 - C_{21}} \\
&\times ((1-\pi_1)\pi_2)^{C_{20}}((1-\pi_1)(1-\pi_2))^{\nu - C_1 - C_{20}}
\end{aligned}
$$

Finding out the full conditional, easy!

How to read the conditional distr. of C_{21} given C_1 and C_2?

$$
\begin{aligned}
[C_1, C_{20}, C_{21}] &= [C_1, C_2 - C_{21}, C_{21}] \\
[C_{21} | C_1, C_2] &\propto \frac{1}{C_{21}!(C_1 - C_{21})!C_{20}!(\nu - C_1 - C_{20})!} \\
&\propto \frac{1}{C_{21}!(C_1 - C_{21})!(C_2 - C_{21})!(\nu - C_1 - C_2 + C_{21})!} \\
&\propto \frac{C_1!(\nu - C_1)!}{C_{21}!(C_1 - C_{21})!(C_2 - C_{21})!(\nu - C_1 - C_2 + C_{21})!} \\
&\propto C_{C_{21}}^{C_1} C_{C_2 - C_{21}}^{\nu - C_1} = \frac{C_{C_{21}}^{C_1} C_{C_2 - C_{21}}^{\nu - C_1}}{C_{C_2}^{\nu}} \\
\text{car } \sum_y C_y^N C_{K-y}^B &= C_K^{N+B} \quad \text{see Newton binomial formula}
\end{aligned}
$$

Traditional sampling without replacement

$$C_1 \sim rbinom(1, \nu, \pi_1)$$
$$C_2 \sim rbinom(1, \nu, \pi_2)$$
$$C_{21}|C_1, C_2 \sim hypergeometrique(1, C_2, C_1, \nu - C_1)$$

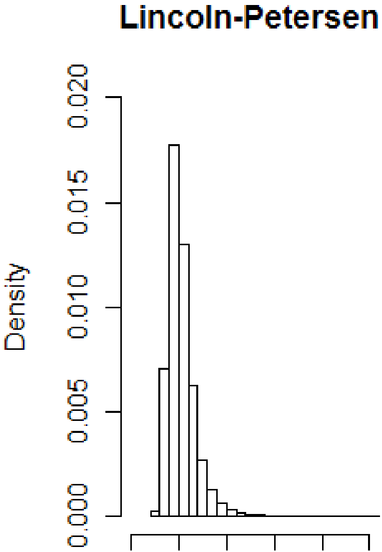
Nota: Y follows a hypergeometric distribution, i.e. one draws K balls without replacement from an urn with N black balls and B white ones

$$[y|K, N, B] = \frac{C_y^N C_{K-y}^B}{C_K^{N+B}}$$
$$\mathbb{E}(Y) = \frac{N}{N+B} K$$
$$\mathbb{E}(\frac{1}{Y+1}) = \frac{N+B+1}{(N+1)(K+1)}$$

Estimates, with recourse to R

```
N=100 theta1=0.3 theta2=0.3 repet=100000
C1=rbinom(repet,N,theta1) M2=rbinom(repet,C1,theta2)
C2=M2+rbinom(repet,N-C1,theta2) Nchap=C1*C2/M2
Nchap2=C1*C1/M2 Nchap3=(C1+1)*(C2+1)/(M2+1)-1
par(mfrow=c(1,2)) hist(Nchap, nc=50,main="Lincoln-Petersen",
xlab="estimateur",xlim=c(0,500),freq=F,ylim=c(0,0.02))
hist(Nchap3,nc=50,main="Schnabel-
Chapman",xlab="estimateur",xlim=c(0,500),freq=F,ylim=c(0,0.02))
```

Competing estimators



Schnabel and the hypergeometric

Lincoln-Petersen estimator

$$\frac{C_1}{\hat{N}} = \hat{\pi}_1 = \hat{\pi}_2 = \frac{C_{21}}{C_2}$$

$$\hat{N} = \frac{C_1 C_2}{C_{21}}$$

Schnabel-Chapman estimator

$$\hat{N} = \frac{(C_1+1)(C_2+1)}{C_{21}+1} - 1$$

$$V(\hat{N}) = \frac{(C_1+1)(C_2+1)}{(C_{21}+1)^2} \frac{(C_1-C_{21})(C_2-C_{21})}{C_{21}+2}$$

Expertise, with a hand from BUGS

model{
C1∼dbin(theta1,nu)
C21∼dbin(theta2,C1)
nonmarques←nu-C1
C20∼dbin(theta2,nonmarques)

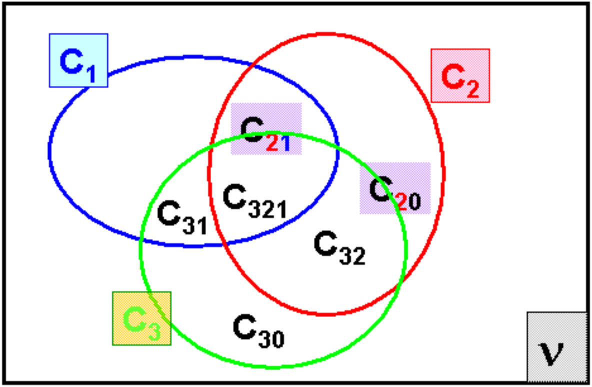
theta∼dbeta(a,b)
theta1←theta theta2←theta theta3←theta
nureal∼dunif(nu_{min}, nu_{max})
nu←trunc(nureal)
a←0.8 b←8-a nu_{min}←50 nu_{max}←250 }
#donnees list(C1=15, C20=16, C21=1)
#inits list(theta=.12, nureal=150) list(C30=11, theta=.25, nureal=100)
list(C30=21, theta=.42, nureal=200)

Full conditional BETA!

Handmade calculus and Raoblackwellisation for π

$$
\begin{aligned}
\pi_1 &= \pi_2 = \pi \\
[\nu, \pi, C_1, C_{20}, C_{21}] &= [C_{20}, C_{21} | C_1, \nu, \pi][C_1 | \nu, \pi][\pi, \nu] \\
&= \frac{\nu! \pi^{C_1 + C_2} (1-\pi)^{2\nu - C_1 - C_2} [\nu][\pi]}{C_{21}!(C_1 - C_{21})! C_{20}! (\nu - C_1 - C_{20})!} \\
[\pi] &\sim dbeta(a, b) \text{ et } [\nu] \text{ discret} \\
[\pi | \nu, C_1, C_{20}, C_{21}] &\sim dbeta(a + C_1 + C_2, b + 2\nu - C_1 - C_2) \\
[\nu | C_1, C_{20}, C_{21}] &= \frac{[\nu, \pi, C_1, C_{20}, C_{21}]}{[\pi | \nu, C_1, C_{20}, C_{21}]} \\
&= \frac{\Gamma(b + 2\nu - C_1 - C_2) \nu! [\nu]}{\Gamma(a + b + 2\nu)(\nu - C_1 - C_{20})!} \times Cste
\end{aligned}
$$

More than two passes: figuring out what happens



European Dippers

State-Space Events

European dipper
1,0,0,0,0,0,0,
1,0,1,0,0,0,0,
1,1,0,0,0,0,0,
1,1,0,1,1,1,0,
1,1,1,1,0,0,0,
1,1,1,1,1,0,0,
1,1,1,1,1,1,0
....
0,1,1,1,1,0,0,
0,1,1,1,1,1,0,
0,1,1,1,1,1,1,
0,1,1,1,1,1,1,
....

State Space Model

Summing up table

Year	Obs: 1981+ Marks	1	2	3	4	5	6
1981	22	11	2	0	0	0	0
1982	60	-	24	1	0	0	0
1983	78	-	-	34	2	0	0
1984	80	-	-	-	45	1	2
1985	88	-	-	-	-	51	0
1986	98	-	-	-	-	-	52

Likelihood from summary statistics

Likelihood

Obs: 0+	1	2	3
proba	ϕp	$\phi^2(1-p)p$	$\phi^3(1-p)^2p$
number	11+24+...	2+1+2+1...	0+0+0+2

See caprecapCC.R et caprecapCC.odc

p = recapture prob.

ϕ= survival prob

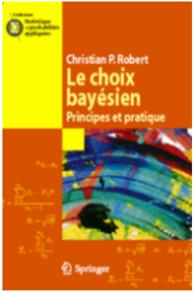
State-space formulation

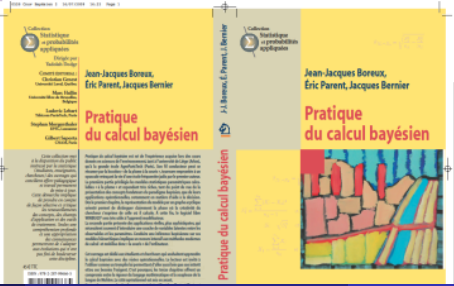
Each individual is in either state 0, 1

- state equation
  ProbaAlive[t] $=\phi$ * Alive[t-1]
  Alive[t] $\sim$ dbern(ProbaAlive[t])
- observation equation
  ProbaSeen[t] $<$
  Seen[t] $\sim$ dbern(ProbaSeen[t])

See caprecapCC-statespace.odc

Bonnes lectures





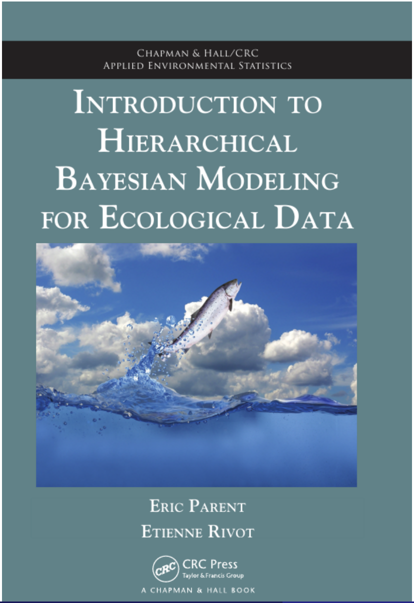
Bonnes lectures

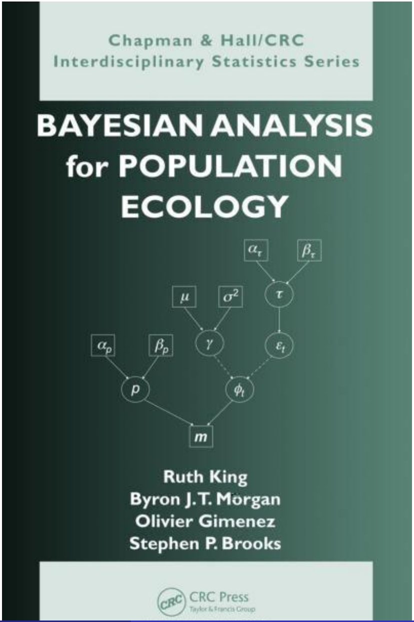
Jeter des ponts

Le monde des expérimentateurs est actuellement imperméable, pour ne pas dire rebelle aux discours tenus par les statisticiens. En gros, c'est un monde des gens qui ont des besoins méthodologiques mais peu d'outils !

A contrario, il existe un monde monodisciplinaire, celui des matheux. Dans ce monde, tous partagent le même intérêt et parfois le même enthousiasme pour disserter sur telle technique mathématique astucieuse, améliorer d'un fifrelin tel estimateur, découvrir et répertorier les cas pathologiques, etc. Bref, ici, c'est souvent le monde des gens qui ont des outils et peu de besoins !

Good to read


CHAPMAN & HALL/CRC
APPLIED ENVIRONMENTAL STATISTICS
INTRODUCTION TO HIERARCHICAL BAYESIAN MODELING FOR ECOLOGICAL DATA
ERIC PARENT
ETIENNE RIVOT
CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK


Chapman & Hall/CRC
Interdisciplinary Statistics Series
BAYESIAN ANALYSIS for POPULATION ECOLOGY
α_τ
β_τ
μ
σ^2
τ
α_p
β_p
γ
ε_t
p
ϕ_t
m
Ruth King
Byron J.T. Morgan
Olivier Gimenez
Stephen P. Brooks
CRC Press