Stat 534: formulae referenced in lecture, week 6: Open population models

Open population: $B_t \neq 0$ and/or $D_t \neq 0$

Notation (many carried over from closed pop statistics), statistics:

t	# capture occasions
M_i	# marks in the population just before sampling at time i
m_i	# marked animals captured at time i
u_i	# unmarked animals captured at time i
n_i	$= m_i + u_i$ # animals captured at time i
R_i	# animals released at time i
	When no loss or death between capture and subsequent release, $R_i = n_i$
	Removal sampling: $R_i = 0$ (open pop models not useful for removal sampling)
r_i	# animals released at time i subsequently caught at least once
z_i	# animals caught before i, not caught at i, caught after i

Parameters (more parameters later for more complicated models):

ϕ_i	Apparent survival probability from time i to $i+1$
B_i	N# births between time i and $i+1$
N_i	# individuals at time i
p_i	capture probability at time i

Apparent survival because can't distinguish death from permanent emigration

Goals:

- Estimate apparent survival rate, ϕ, or ϕ_i
- Estimate recruitment = births during each interval
- Estimate population size (usually much less important)

Big change from closed pop: How many marked individuals in the population?

- Closed population: all individuals survive, so can count # marks, $M_i = \sum u_i$
- Open population: a marked individual may die → fewer marks in the population
- Births are born without marks, so birth has no effect on # marked individuals

Picture of a capture history:

Start by assuming we have an estimate of $\widehat{M}_i$ (# marks in population)

Population size, N_i:

- LP estimate:
  - Assume sample proportion of marked individuals = pop. proportion
  - $m_i/n_i = M_i/N_i$
  - $\widehat{N}_i = n_i \widehat{M}_i / m_i$

Survival, ϕ_i: consider marked animals only

- Just after capture and release at time i, # marked individuals = $R_i + (\widehat{M}_i - m_i)$

- at time $i+1$, estimated # marked animals is $\widehat{M}_{i+1}$

$$\begin{aligned}
\phi_1 &= \frac{\widehat{M}_2}{R_1} \\
\phi_i, i = 2, 3, \cdots t-2 &= \frac{\widehat{M}_{i+1}}{\widehat{M}_i - m_i + R_i} \\
\phi_{t-1} & \quad \text{unknown because can't estimate } M_t \\
\phi_t & \quad \text{unknown because can't estimate } M_{t+1}(after end of study)
\end{aligned}$$

Recruitment, B_i: = # at N_{i+1} - expected # survivors from N_i

$$\widehat{B}_i = \widehat{N}_{i+1} - \hat{\phi}_i \left[\widehat{N}_i - (n_i - R_i) \right]$$

- if no loss on capture, $n_i = R_i$, so $(n_i - R_i) = 0$, $\widehat{B}_i = \widehat{N}_{i+1} - \hat{\phi}_i \widehat{N}_i$

Capture probability, p_i

- proportion of marked animals caught, $\hat{p}_i = m_i / \widehat{M}_i$
- or proportion of population caught, $\hat{p}_i = n_i / \widehat{N}_i$

Estimating M_i:

Four capture history fragments, want to estimate $\widehat{M}_i$ at time i

ID	Sometime in past	Time i ⇓	Sometime in future $i+1$	$i+2$	$i+3$
1	Y	Y	N	Y	N
2	Y	Y	N	N	N
3	Y	N	N	N	Y
4	Y	N	N	N	N

- 1: obviously counts in $\widehat{M}_i$ because caught at i
- 2: ditto, but alive and frequently missed after i, or died at some time along the way
- 3: counts in $\widehat{M}_i$ because known alive at i (caught later)
- 4: don't know whether in $\widehat{M}_i$, may be alive and uncaught at i; may have died before i

Jolly's and Seber's (separately) incredible insight:

- Two groups of individuals: those seen at i and those not seen at i
- Consider the probability at an individual will be seen sometime after i

Seen at i	$\frac{r_i}{R_i}$	proportion of those seen at i that are seen later
Not seen at i	$\frac{z_i}{M_i - m_i}$	proportion of those not seen at i that are seen later

- Notation reminder:
  - r_i: # animals released at i captured at least once subsequently
  - z_i: # captured before i, not at i, and at least once subsequently
- Should be the same for both groups, so set them equal and can solve for M_i

$$\widehat{M}_i = m_i + \frac{R_i z_i}{r_i}$$

- Only defined for $i = 2, 3, \cdots, t-1$
  - $M_1 = 0$ at start of study
  - M_t not defined - no captures after last time, t

Statistical note:

- $\widehat{M}_i$ is a ratio of random variables
- ratios of random variables often provide biased estimates

Bias corrections:

$$\begin{aligned}\widetilde{M}_i &= m_i + \frac{(R_i+1)z_i}{r_i+1} \\ \widetilde{N}_i &= \frac{(n_i+1)\widetilde{M}_i}{m_i+1}\end{aligned}$$

- All approximately unbiased
- The +1 terms have most impact when other component, e.g., R_i, is small

Turning into a likelihood: likelihood has 3 components

1. dealing with # unmarked caught each time and their capture probabilities
2. probability of loss on capture
3. capture and survival probabilities for marked individuals

Cormack-Jolly-Seber

- only consider marked animals
  - ignore anyone's fate before they get caught the first time
  - Can't estimate population size
  - Can estimate # marked individuals, but not usually interesting
  - Can't estimate births (born without marks)
- Condition on first capture time
- follow capture history after they get marked
- Can estimate apparent survival and capture probabilities
- 2 parameters for most capture occasions:
  - p_i: P[capture on occasion i]
  - ϕ_i: P[survive from i to $i+1$], i.e. doesn't die, doesn't emigrate
- likelihood is a function of survival and capture probabilities
- N_i is not part of the likelihood

"Easy" to write the likelihood from the probability of a capture history.
3 examples:

ID	Capture history	Probability
1	N Y N Y N Y	$\phi_2(1-p_3)\phi_3 p_4 \phi_4(1-p_5)\phi_5 p_6$
2	N Y N N Y N	$\phi_2(1-p_3)\phi_3(1-p_4)\phi_4 p_5 \chi_5$
3	N Y Y Y N N	$\phi_1 p_2 \phi_2 p_3 \phi_3 p_4 \chi_4$

Explanations of capture probabilities:

- Why no p_1 or ϕ_1? Before first capture
- Why no p_2? Because conditioning on first capture, so P[caught at time 2 | caught at time 2] = 1
- If an individual caught in time 1, probability would include ϕ_1 and either p_2 or $(1-p_2)$
- χ_j = P[not seen after time j], could have died or not caught
  - χ_5 = P[die 5-6] + P [survive 5-6]*P[not seen in 6]
  - $\chi_5 = 1 - \phi_5 + \phi_5(1-p_6)$
- $\chi_t = 1$ because no one seen after the last sampling occasiont
- χ_4 = P[not seen after 4] has more possible fates:
  P[die 4-5] + P[survive 4-5, not seen in 5, died 5-6]
  + P[survive 4-5, not seen in 5, survived 5-6, not seen in 6]
- Can write a recursion (start at time $j = t-1$)

$$\chi_j = 1 - \phi_j + \phi_j(1-p_{j+1})\chi_{j+1}$$

Notice that you can't estimate all capture probabilities and all survival probabilities

- p_1: because conditioning on first capture, numerically = 1, but not interesting
- ϕ_6: = P[survive from 6 to 7], goes outside study period
- ϕ_5: only term in likelihood is $\phi_5 p_6$
  - Can't separately estimate ϕ_5 or p_6, only their product
  - Data can't distinguished died in penultimate period from not caught in last period
  - because no opportunity to observe 'caught after time 6'
  - could impose some constraint, e.g., $p_5 = p_6$, but arbitrary and usually not done
- Why does ϕ_5 always occur with p_6 in the 6 period likelihood?

- Fragments of capture histories for last two period for an individual caught in time 5
  - Y Y: $\phi_5 p_6$
  - Y N: $1 - \phi_5 + \phi_5(1 - p_6)$: looks like this breaks the "always $\phi_5 p_6$, but:
  - Y N: $= 1 - \phi_5 + \phi_5 - \phi_5 p_6 = 1 - \phi_5 p_6$

Useful insight:

- The capture probability factors into independent components for each release
- Consider this capture history: N Y N Y N Y
- Capture probability = $\phi_2(1 - p_3)\phi_3 p_4 \phi_4(1 - p_5)\phi_5 p_6$
- Two independent components:
  - from time 2 - 4, conditioning on capture in time 2: $\phi_2(1 - p_3)\phi_3 p_4$
  - from time 4 - 6, conditioning on capture in time 4: $\phi_4(1 - p_5)\phi_5 p_6$
- The likelihood is the same if this is one individual,
- or two individuals, one with N Y N Y (then not released) and one with N N N Y N Y
- Supports Nichols's definition of n (e.g., for AICc or BIC computation) as # releases