Stat 534: formulae referenced in lecture, week 5, part 2:
Overparameterized models and models accounting for individual heterogeneity
Corrected, 30 Sept 21

Mtb: both behavioural response and time-dependent capture probabilities

- Our first encounter with an overparameterized mark-recapture model
- For 3 occasions: 6 parameters, 3 p's, 2 c's (no c_1) and 1 N
- Let's work out the capture probabilities

Time 1	2	3	# animals	probability
Y	Y	Y	n_{111}	$p_1 c_2 c_3$
Y	Y	N	n_{110}	$p_1 c_2 (1 - c_3)$
Y	N	Y	n_{101}	$p_1 (1 - c_2) c_3$
Y	N	N	n_{100}	$p_1 (1 - c_2)(1 - c_3)$
N	Y	Y	n_{011}	$(1 - p_1) p_2 c_3$
N	Y	N	n_{010}	$(1 - p_1) p_2 (1 - c_3)$
N	N	Y	n_{001}	$(1 - p_1)(1 - p_2) p_3$
N	N	N	n_{000}	$(1 - p_1)(1 - p_2)(1 - p_3)$

- Nothing looks amiss here, so let's count sufficient statistics
- based on the patterns for Mt and Mb, you would expect:
  - n_i: number caught at time i
  - m_i: number marked caught at time i
  - M_{t+1}: total number of unique individuals seen
- Looks like 6 sufficient statistics, enough to estimate 6 parameters
- But: M_{t+1} is not "new" information. It can be computed from n_i and m_i
  - # 1st seen at time 1 + # first seen at time 2 + # first seen at time 3
  - $= n_1 + (n_2 - m_2) + (n_3 - m_3)$
- so no unique solution! Many sets of parameters have the same lnL
- Sometimes can spot redundancy by examining the capture history probabilities
- Have a problem when two parameters only appear multiplied or added together, e.g. $\theta_1 \theta_2$ or $\theta_1 + \theta_2$
- Numerical assessments:

- look at the rank of the negative Hessian matrix (= # non-zero eigenvalues)
- often, use very small (e.g., < 0.005) because of numerical issues computing the Hessian
- For one particular data set, the eigenvalues were:
  44.1339 25.0000 20.8004 18.6077 0.0022 0.0015
  Two parameters have very large variances, so can not be identified
- or look for massive standard errors for one or more parameters

Solutions for overparameterized models:

- Simplify the model, e.g. reparameterize $\theta_1\theta_2$ as a single parameter
- Put constraints on the parameters
- E.g., for Mtb you could:
  - Make time effects follow a logistic regression

$$\text{logit } p_i = \log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 i$$

  - Connect recapture and capture probabilities, also usually by a logistic

$$\text{logit } c_i = \nu + \text{logit } p_i$$

Model Mh: allows each of the N individuals to have a different p_i

- The capture history for five **individuals**:

Time 1	2	3	# animals	probability
Y	Y	Y	1	p_1^3
Y	Y	Y	1	p_2^3
Y	Y	N	1	$p_3^2(1-p_3)$
N	N	N	1	$(1-p_4)^3$
N	N	N	1	$(1-p_5)^3$

- Sufficient statistics are the # captures and # missed for each of the M_{t+1} animals seen at least once
- There are M_{t+1} sufficient statistics, but N+1 parameters
- Can't estimate N without modeling p_i or using a completely different approach (coverage)

Modeling p_i using a Beta(a,b) distribution

- $f(p|a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p^{a-1}(1-p)^{b-1}$
  - $\Gamma(x)$ is the gamma function
  - Uniform(0,1) is Beta(1,1)
  - Hand drawn pictures
  - Key is that all are unimodal except when peaks are at 0 or 1
- mean: $\mu_p = \text{E } p = \frac{a}{a+b}$
- variance: $\text{Var } p = \frac{ab}{(a+b)^2(a+b+1)} = \mu_p(1-\mu_p)\frac{1}{a+b+1}$
- The approach:
  - Estimate a and b from the capture histories of the M_{t+1} observed individuals
  - Assume all N individuals have capture probabilities from that Beta(a, b) distribution
  - Provides sufficient information to estimate N
- Ken Burnham worked on this model extensively in the 1980's
- Great idea in theory, didn't work in practice
- Problem is that capture probabilities are more complicated than can be modeled by a Beta distribution

Pledger heterogeneity models (2000)

- For now, focus on Mh: no time or behavioural effects
- p_i = probability individual i is captured on an occasion
- P[capture history | p_i] = P[capture history | p_i] = $\prod_{j=1}^{t} p_i^{X_{ij}}(1-p_i)^{(1-X_{ij})}$
- If don't know p_i, then want the marginal probability

$$\text{P[capture history]} = \int_{p_i=0}^{1} f(\text{capture history} \mid p_i) f(p_i)\, d\, p_i$$

- Burnham's beta model used this idea with a beta distribution for $f(p_i)$
  - beta and binomial distributions "play nicely" together
  - the integral has a analytical solution
- Shirley Pledger proposed a more flexible distribution for $f(p_i)$

- finite mixture model: p_i has one of two values (could be 3 or more but 2 often sufficient)

$$f(p_i) = \begin{cases} \pi_1 & \text{with probability } f_1 \\ \pi_2 & \text{with probability } f_2 = 1 - f_1 \end{cases}$$

- So, the marginal distribution is a sum, not an integral, because p_i has only 2 values:

$$\text{P[capture history]} = \sum_{a=1}^{2} f_a \ \text{P[capture history}|p_a]$$

- for 2 components: 3 parameters for capture probability: f_1, π_1, π_2.
  - More parameters if more than 2 mixture components,
  - e.g., for 3 components: f_1, f_2, π_1, π_2, π_3
- Likelihood is (for 2 components):

$$L(N, f_a, \pi_1, \pi_2 \mid \{X_{ij}\}) = \frac{N!}{\text{constant} \times (N - M_{t+1})!} \prod_{i=1}^{N} \left[\sum_{a=1}^{2} f_a \ \text{P[capture history}|p_a] \right]$$

- lnL is not simple - includes log sum(stuff)
- Can generalize to include time and behavioural effects

Huggins model: relate p_i to one or more covariates, X_i

- Logistic regression model: $\text{logit}(p_i) = \log \frac{p_i}{1-p_i} = \beta_0 + \beta_1 X_i$

$$p_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_i)}}$$

- Not fit using the "usual" logistic regression algorithms
  - Only have data for individuals seen at least once - missing many of the 0's
  - Need to model P[capture history | captured at least once]

$$\text{P[capture history} \mid \text{captured at least once]} = \frac{\text{P[(capture history) \& (captured at least once)]}}{\text{P[captured at least once]}}$$

- P[capture history and captured at least once] = product of appropriate p or p and c terms (usual expression)
- Define p_{ij} = P[capture individual i on occasion j]
- P[captured at least once] = 1 - P[never captured] $= 1 - \prod_{j=1}^{k}(1 - p_{ij})$

- Huggins incorporated time and behavioural responses by adding
  - a unique value for each time
  - First captures: note time-specific intercept

$$\text{logit } p_{ij} = \beta_{0j} + \beta_1 X_i$$

  - c (recapture probs) have a fixed relationship to p (first capture probs)
  - Recaptures: the p_{ij} model + a consistent trap-happy or trap-shy effect, ν

$$\text{logit } c_{ij} = \nu + \beta_{0j} + \beta_1 X_i = \nu + \text{logit } p_{ij}$$

- Provides a simple way to model c and p
- What about estimating N?
  - N not in the likelihood
  - need some other way to estimate it

Huggins model: estimating N

- Add up 1/P[seen at least once] for each observed animal
- (Horvitz-Thompson estimator: Population Total $Y = \sum_{i=1}^{n} Y_i/\pi_i$)

$$\widehat{N} = \sum_{i=1}^{M_{t+1}} \frac{1}{1 - \prod_{j=1}^{t}(1 - \hat{p}_{ij})}$$

- Can get a variance estimator (complicated, not illuminating)

Is Huggins $\hat{N}$ a reasonable estimator?

- Prefer unbiased estimates.
  - $\hat{N}$ is a random variable.
  - Would like $\text{E } \hat{N} = N$
  - i.e., sometimes $\hat{N}$ is too high, sometimes too low, but on average spot on
- Is $\hat{N}_{Huggins}$ unbiased?
  - Yes, if p_{ij} is known precisely
  - Define $p_i(\theta) = \text{P}[\text{individual } i \text{ captured at least once} \mid \text{parameters } \theta]$
  - Define $C_i = \begin{cases} 0 & \text{when individual not captured} \\ 1 & \text{when individual was captured} \end{cases}$

$$\text{E } \hat{N}_{Huggins} = E \sum_{i=1}^{M_{t+1}} \frac{1}{p_i(\theta)} = E \sum_{i=1}^{N} \frac{C_i}{p_i(\theta)} = \sum_{i=1}^{N} \frac{\text{E } C_i}{p_i(\theta)} = \sum_{i=1}^{N} \frac{p_i(\theta)}{p_i(\theta)} = N$$

- 3rd step requires known $p_i(\theta)$
- In practice, $p_i(\theta)$ is estimated, because coefficients in the models for p_{ij} and c_{ij} are estimated
- bias is small so long as sufficient data to provide good estimates of $p_i(\theta)$

Sample Coverage (Chao heterogeneity estimators)

- Coverage: proportion of total capture probability associated with observed individuals

$$C = \sum_{i=1}^{M_{t+1}} \frac{p_i \text{ I(seen at least once)}}{\sum_{i=1}^{N} p_i}$$

- Explanation / example in hand-written notes
- Notation:
  - $\overline{p}$ mean capture probability in the population
  - γ c.v. of capture probabilities in the population
  - n total # captures
  - f_i # number of individuals seen i times, e.g.:
  - f_1 # number seen only once
  - f_2 # number seen twice, etc.
- Assume a particular model for the data (details in Chao & Lee 1992, 1994, not important)
  use to derive expected values

$$\frac{\text{E } M_{t+1}}{\text{E } C} \cong N - \frac{n\,(1-\overline{p})^{n-1}}{\text{E } C}\gamma^2$$

- Need to estimate C, $\overline{p}$, and γ
- $\overline{p}$:
  - Have p_i for observed individuals, but average is biased
  - More likely to observe individuals with large p_i
  - $\overline{p}$ in the population can be estimated from f_1: # individuals seen only once

$$N \cong \frac{\text{E } M_{t+1}}{\text{E } C} + \frac{\text{E } f_1}{\text{E } C}\gamma^2$$

- γ: sample cv not so obviously biased
- C: an old estimator, from the 1950's

$$\hat{C} = 1 - \frac{f_1}{\sum_{i=1} ti\, f_i}$$

- There is a bias corrected estimate, but really want unbiased estimates of $1/\hat{C}$ and $\gamma^2/\hat{C}$

- Putting the pieces together: applying the "plug-in" principle

$$\hat{N} = \frac{M_{t+1}}{\hat{C}} + \frac{f_1}{\hat{C}}\hat{\gamma}^2$$