Stat 534: formulae referenced in lecture, week 5:
Otis closed population models

Aside on coefficient of variation (cv) and log scale parameters

- Reminder: cv describes relative variability
  - Usually used to describe data: cv = sd / mean
  - Can also be used to describe relative variability in an estimate: $\text{cv}_{estimate} = \text{se}_{estimate}$ / estimate
- close relationship between sd of a log-scale estimate and cv of that estimate
- Consider an estimate of a log-scale parameter, e.g., $\gamma = \log\theta$.
  You may have noticed that I:
  - have exponentiated γ to estimate θ, $\exp\gamma = \exp\log\theta = \theta$
  - have exponentiated lower and upper confidence limits for γ to get the ci for θ
  - **have not** exponentiated the se of γ
- General approximation: Var $\log\theta \cong$ Var $\theta/(\theta)^2$
  - So cv of $\theta = \sqrt{\text{Var }\theta}/\theta \cong \text{sd } \log\theta$
- When $\gamma = \log\theta \sim N(\mu_l, \sigma_l^2)$, θ has a log normal distribution
  - E $\theta = \exp(\mu_l + \sigma_l^2/2)$
  - median $\theta = \exp\mu_l$
  - so mean of $\theta >$ median of θ. By how much depends on σ_l^2
  - Var $\theta = [\exp(\sigma_l^2) - 1]\exp(2\mu_l + \sigma_l^2) = (\text{E }\theta)^2\,[\exp(\sigma_l^2) - 1]$
  - cv $\theta = \sqrt{\text{Var }\theta/(E\theta)^2} = \sqrt{\exp(\sigma_l^2) - 1}$
- Connection between the two formulae
  - When σ_l^2 close to 0, $\exp(\sigma_l^2) \cong 1 + \sigma_l^2$ so cv $\theta \cong \sqrt{1 + \sigma_l^2 - 1} = \sigma_l$.
  - Normal formula relies on that normality, but works for any σ_l^2
  - Approximation works for any distribution, best for small σ_l^2

Redpoll data example ($n_1 = 13$, $n_2 = 45$, $m_2 = 9$)

Data set	$\hat{N}$	sd $\hat{N}$	cv $\hat{N}$	Var log $\hat{N}$	sd log $\hat{N}$	normal
redpoll	90.14	20.13	22.3%	0.0514	0.226	23.0%
2x	184.2	29.7	16.1%	0.0261	0.161	16.2 %
25x	2333	107.2	4.6%	0.0021	0.046	4.6%

M0: constant capture probability

- Capture histories and their probabilities

Time				
1	2	3	# animals	probability
Y	Y	Y	n_{111}	p^3
Y	Y	N	n_{110}	$p^2(1-p)$
Y	N	Y	n_{101}	$p^2(1-p)$
Y	N	N	n_{100}	$p(1-p)^2$
N	Y	Y	n_{011}	$p^2(1-p)$
N	Y	N	n_{010}	$p(1-p)^2$
N	N	Y	n_{001}	$p(1-p)^2$
N	N	N	n_{000}	$(1-p)^3$

- When you write out the log likelihood contributions from each unique capture history and combine terms, you get the log likelihood at the end of the week 4 notes

Mt: capture probability depends on occasion

- General model: unique capture probability for each occasion
- Sampling design and associated models often given names based on early investigators
- Study with multiple sampling occasions often called a Schnabel census
  Schnabel (1938) The estimation of the total fish population of a lake. *Am. Math. Monthly* 45:348-352
- $t+1$ parameters: $p_1,\ p_2,\ \cdots\ p_t,\ N$

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

- Writing out each contribution to lnL and combining terms gives:

$$\text{lnL}(p_1, p_2, \cdots p_t, N \mid n_1, n_2, \cdots, n_t, M_{t+1}) = \log[N!] - \log[(N-M_{t+1})!] - \mathit{constant}$$
$$+ n_1 \log p_1 + (N-n_1)\log(1-p_1) + n_2 \log p_2 + (N-n_2)\log(1-p_2) + \cdots$$
$$+ n_t \log p_t + (N-n_t)\log(1-p_t)$$

- Notation:

n_i	total # caught on occasion i
M_{t+1}	# individuals seen at least once = # tags in population **after** the last occasion

- mle of N requires finding the root of an $t-1$'th degree polynomial
  - Easy for $t = 2 \Rightarrow$ LP estimator
  - Now, numerical maximization usually used for $t > 2$
- Other data models (e.g., versions of binomial models or hypergeometric models) lead to other estimators
- Note that $\hat{p}_i = n_i/\hat{N}$, if you have an estimate of N
  - So easy to optimize profile likelihood $\text{lnL}(p_1,\ p_2, \cdots p_t \mid N)$

Mb: behavioural heterogeneity

- Capture probability different for 1st captures and subsequent captures
- "trap-happy" or "trap-shy" behaviours
- Notation:
  p P[capture | never captured before]
  c P[capture | captured already, at least once]
- Both p and c assumed constant over time
  - Model Mtb generalizes this to time-dependent p and c
- 3 parameters, no matter how many capture occasions: p, c, N
- Capture histories and their probabilities

	Time			
1	2	3	# animals	probability
Y	Y	Y	n_{111}	pc^2
Y	Y	N	n_{110}	$pc(1-c)$
Y	N	Y	n_{101}	$p(1-c)c$
Y	N	N	n_{100}	$p(1-c)^2$
N	Y	Y	n_{011}	$(1-p)pc$
N	Y	N	n_{010}	$(1-p)p(1-c)$
N	N	Y	n_{001}	$(1-p)^2p$
N	N	N	n_{000}	$(1-p)^3$

- Notation:

n_i:	# number of individuals caught at time i
m_i	# number marked individuals caught at time i
$m_.$	total # times marked individuals were captured, $m_. = \sum_i m_i$
M_i	# marked individuals **in the population** at time i (before start trapping)
$M_.$	$\sum_{i=1}^{t} M_i$

Mb: log likelihood function

- After combining terms in the multinomial log likelihood, you will see that:
  - p occurs M_{t+1} times (each animal first seen only once)
  - c occurs $m_.$ times (total number of captures of already marked animals)
  - $1-c$ occurs $M_. - m_.$ times (number of "capturable" marked animals that weren't captured)
  - $1-p$ occurs $tN - M_{t+1} - M_.$ times (by difference, hard to intuit)
- Hence the sufficient statistics are M_{t+1}, $M_.$, $m_.$
- So we can estimate 3 parameters
- The log likelihood is:

$$\begin{aligned}\text{lnL}(p, c, N \mid n_1, n_2, \cdots, M_{t+1}, M_., m_.) &= \log [N!] - \log [(N - M_{t+1})!] - \textit{constant} \\ &+ M_{t+1} \log p + (tN - M_{t+1} - M_.) \log(1-p) + m_. \log c + (M_. - m_.) \log(1-c)\end{aligned}$$

- Differentiating and solving gives:

$$\begin{aligned}\hat{p} &= \frac{M_{t+1}}{t\hat{N} - M_.} \\ \hat{c} &= \frac{m_.}{M_.}\end{aligned}$$

- $tN - M_.$ is total # not yet caught occasions
  - Y Y Y = 0
  - Y Y N = 0
  - N Y Y = 1
  - N N Y = 2
- $\hat{N}$ doesn't depend on $m_.$
  - So second and subsequent captures provide no information about p (makes sense) or N (surprising)
  - c can be any value, without sacrificing information about N or p

Removal sampling:

- Don't return marked animals immediately, so $c = 0$
- Example

Occasion	# caught
1	260
2	141
3	97
4	50

- Don't have to continue until you fail to catch more!
- Use model Mb with $m_{.} = 0$ so $\hat{c} = 0$
  - Could use Mtb if p is not constant

Choosing a model:

- $\text{AIC} = -2 \text{ lnL} + 2k$
  - k is the number of parameters in the model
- This is an asymptotic result
- Small-sample corrected AIC: $\text{AICc} = -2 \text{ lnL} + 2k + \frac{2k(k+1)}{n-k+1}$
  - Originally developed in the time-series literature:
    Hurvich and Tsai 1989, Biometrika 76:297-307
  - n = # observations
  - "useful when n/k < 40"
- What is n for mark recapture data?
  - Best answer (so far): total # releases (Nichols)
  - so an individual captured (and released) twice adds 2 to n
- $\text{BIC} = -2 \text{ lnL} + k \log n$
  - more penalty per parameter when $n \geq 8$
  - commonly used outside of wildlife
  - wildlife prefers AIC or AICc
- Same difficulty: what is n?
  - I don't believe anyone has investigated properties of the n = # releases suggestion

Example: Reid deermice data, *Peromyscus maniculatus*, 6 days, 99 traps per day, $n = 133$

Model	k	lnL	AIC	AICc	BIC
M0	2	-57.635	119.27	119.36	125.0
Mt	7	-47.405	104.81	105.76	129.0
Mb	3	-43.422	92.84	93.03	101.5

Model selection / model averaging:

- Made up data: $t = 5$, minimum known alive $= M_{t+1} = 30$

model	k	$\hat{N}$	$\widehat{\text{Var}}\ \hat{N} \mid model$	lnL	AIC	Δ AIC	$\exp(-\Delta/2)$	weight
M0	2	50	30	-8.00	20.0	4.5	0.105	0.042
Mt	6	70	40	-2.25	16.5	1.0	0.606	0.243
Mb	3	90	60	-4.75	15.5	0	1.0	0.401
Mtb	7	80	60	-1.00	16.0	0.5	0.779	0.312

- Which model?
  - Burnham and Anderson, classic advice:
    - Δ AIC < 2 model relatively well supported by data
  - B&A, more recent advice:
    - Δ AIC < 4 model relatively well supported by data
  - Relationship between AIC choice and p-values, 2 nested models, H0: simpler model, Ha: add 1 parameter
    - Choose model with smaller AIC: Δ AIC $= 0 \Rightarrow$ LRT p-value $= 0.16$
    - Consider 2 models with Δ AIC $= 0$ and $= 1.84 \Rightarrow$ LRT p-value $= 0.05$
    - Consider 2 models with Δ AIC $= 0$ and $= 2.00 \Rightarrow$ LRT p-value $= 0.046$
    - Consider 2 models with Δ AIC $= 0$ and $= 4.00 \Rightarrow$ LRT p-value $= 0.014$
  - classic & more recent: Δ AIC > 10 model not well supported by data
- Mb has smallest AIC: If choose that, $\hat{N} = 90$, se $\hat{N} = \sqrt{60} = 7.7$

Model averaging:

- Combine information from all fitted models, more emphasis on estimates from better fitting models
- Bayesian MA
  - Solid theoretical justification
  - Requires a deep dive into Bayesian methods

- Frequentist MA: Start with a list of fitted models
- In wildlife, AIC or AICc used to estimate model weights
  - Need AIC statistics for each model
  - For each model, calculate change in AIC from the best = ΔAIC_i for model i
    - Include best model, for which $\Delta AIC_i = 0$
  - Calculate unnormalized weights for each model $w_i^* = \exp(-\Delta AIC_i/2)$
    - These formulae are for $\Delta AIC_i \geq 0$
    - Use $\exp(\Delta AIC_i/2)$ if $\Delta AIC_i \leq 0$
  - normalize the weights to sum to 1: $w_i = w_i^* / \sum w_i^*$
- MA estimate of $\hat{\theta}_w = \sum_i w_i \hat{\theta}_i$
  - w_i is the weight for model i
  - $\hat{\theta}_i$ is the estimate from model i

Inference on model averaged estimates

- Not an easy problem in the frequentist world, but see mata CIs (next week)
  - Multiple suggested solutions
- Bayesian MA avoids many of the frequentist problems
  - But introduces a new one: what are the prior probabilities for each model?
  - Are simpler models more likely? (i.e., have higher prior probability)
- Active research area, here are current simple approaches
  - Fletcher, D., 2018, Model Averaging, Springer, reviews current approaches

Standard error of MA estimate

- Quantities needed:

Var $\hat{\theta}_i \mid M_i$:	estimated variance of $\hat{\theta}$ from model M_i
$\hat{\theta}_i$:	estimate of θ from model M_i
$\hat{\theta}_w$:	weighted MA estimate of θ

  - Buckland et al. (1997) estimator: se $\hat{\theta}_w = \sum w_i \sqrt{\text{Var}(\hat{\theta} \mid M_i) + (\hat{\theta}_i - \hat{\theta}_w)^2}$
  - "Revised formula": se $\hat{\theta}_w = \sqrt{\sum w_i \left[\text{Var}(\hat{\theta} \mid M_i) + (\hat{\theta}_i - \hat{\theta}_w)^2\right]}$

- Notes:
  - squared bias added to model-specific variance: accounts for estimates far from overall average
  - equivalent to statistics Mean-Squared Error = Var + $(\text{bias})^2$ (not ANOVA MSE)
  - Buckland averages $\sqrt{MSE}$, Revised averages MSE
  - Averaging variance or MSE more typical, Buckland an ad hoc solution to correlated estimates
  - Revised Var always $\geq$ Buckland (Cauchy-Schwarz inequality)
- For made-up data example:
  - Assume Mb is the correct model: $\hat{N} = 90$, se $\hat{N} = \sqrt{60} = 7.7$
  - MA estimate: $\hat{\theta}_w = 80.3$
  - Buckland: se $\hat{\theta}_w = 11.6$
  - Revised: se $\hat{\theta}_w = 12.5$
- A complication: both se formulae assume weights are known values
- They are random variables
  - Introduces additional uncertainty in se $\hat{\theta}_w$
  - and a nasty potential for bias
- Imagine that each model gives an unbiased estimate of $\hat{\theta}_i$
- i.e., E $\hat{\theta}_i = \theta$
- When weights are fixed values, E $\sum w_i \hat{\theta}_i$ is unbiased, $= \sum w_i E\hat{\theta}_i = \theta$
- When weights are random, E $\sum w_i \hat{\theta}_i = \sum (\text{E } w_i) \left(\text{E } \hat{\theta}_i\right) + \text{Cov } w_i \hat{\theta}_i$
- unbiased only when no correlation between weights and estimates

Confidence interval for MA estimate

- Even harder problem for frequentist inference
  - mle theory $\Rightarrow \hat{\theta}_i$ has an asymptotic normal distribution
  - distribution of $\hat{\theta}_w$ is a mixture of normal distributions
- model-averaged-tail-area (mata) confidence intervals