Appendix A. Likelihood functions, negative log-likelihood functions and gradients of the negative log-likelihood function for all likelihoods.

**Notation:**
The data consist of $N$ point counts conducted with $x_{nj} =$ the number of different birds first detected in the $j$th temporal interval of the $n$th count. We define $x_{n.} \equiv \sum_{j=1}^{3} x_{nj}$. For counts, the full data (over all $1 \leq n \leq N$) are denoted $\{x_{n.}\}$, or simply $x_{n.}$ when the full set of data is obvious from context (i.e. in likelihood equations). Similarly, for the removal mixtures (over all $1 \leq n \leq N$ and over all $1 \leq j \leq 3$) the data are denoted $\{x_{nj}\}$ or simply $x_{nj}$. The likelihood is denoted $l$, and the negative log-likelihood is denoted $-L = -\log(l)$.

The gradients of the negative log-likelihood are also provided below. They greatly assist with convergence of these functions, especially the removal mixtures. They must also be multiplied (i.e. using the chain rule) by the gradients of the inverse link functions with respect to the linear coefficients. Letting $v_i$ represent the $i$th design variable and $\beta_i$ the associated design parameter, we write the linear predictor: $\sum_i \beta_i v_i$, and, by convention, $v_1 = 1$ (the design variable associated with the intercept). For the two link functions employed in this manuscript (log and logit) these gradients are given below. An example of the product is given in gradients of the negative log-likelihood for the Poisson count model.

For parameters for which the log-link is used $(k, \mu, \lambda)$:
$$\theta = \exp \left( \sum_i \beta_i v_i \right) \quad \text{and} \quad \frac{d\theta}{d\beta_i} = v_i \theta$$

For the NB distribution with $r = 1/k$:
$$\theta = \exp \left( -\sum_i \beta_i v_i \right) \quad \text{and} \quad \frac{d\theta}{d\beta_i} = -v_i \theta$$

For parameters for which the logit link is used $(c, q, \phi)$:
$$\theta = \left( 1 + \exp \left( -\sum_i \beta_i v_i \right) \right)^{-1} \quad \text{and} \quad \frac{d\theta}{d\beta_i} = v_i \theta (1 - \theta)$$
I. Poisson Counts

\[ l(\lambda | x_{n*}) \propto \prod_{n}^{N} \frac{e^{-\lambda} \lambda^{x_{n*}}}{\Gamma(x_{n*} + 1)}. \]

\[-L = \sum_{n} \left( \lambda + \log(\Gamma(x_{n*} + 1)) - x_{n*} \log(\lambda) \right).\]

\[-\frac{dL}{d\lambda} = \sum_{n} \left( 1 - \frac{x_{n*}}{\lambda} \right) \]

\[-\frac{dL}{d\beta} = \sum_{n} \beta_{n} (\lambda - x_{n*}) \]

II. ZIP Counts

For \( x_{n*} = 0 \):

\[ l(\lambda, \phi | \{x_{n*}\}) \propto \prod_{n: x_{n*} = 0} \left( \phi + (1 - \phi)e^{-\lambda} \right) \]

\[-L = -\sum_{n: x_{n*} = 0} \log(\phi + (1 - \phi)e^{-\lambda}) \]

\[-\frac{dL}{d\phi} = -\sum_{n: x_{n*} = 0} \left( \phi + (1 - \phi)e^{-\lambda} \right)^{-1} (1 - e^{-\lambda}) \]

\[-\frac{dL}{d\lambda} = \sum_{n: x_{n*} = 0} \left( \phi + (1 - \phi)e^{-\lambda} \right)^{-1} (1 - \phi)e^{-\lambda} \]

For \( x_{n*} > 0 \):

\[ l(\lambda, \phi | \{x_{n*}\}) \propto \prod_{n: x_{n*} > 0} \left( 1 - \phi \right) \frac{e^{-\lambda} \lambda^{x_{n*}}}{x_{n*}!} \]

\[-L = \sum_{n: x_{n*} > 0} \left( -\log(1 - \phi) + \lambda - x_{n*} \log(\lambda) + \log(\Gamma(x_{n*} + 1)) \right) \]

\[-\frac{dL}{d\lambda} = \sum_{n: x_{n*} > 0} \left( 1 - \frac{x_{n*}}{\lambda} \right) \]

\[-\frac{dL}{d\phi} = \sum_{n: x_{n*} > 0} \frac{1}{1 - \phi} \]

\[-\frac{dL}{d\beta} = \sum_{n: x_{n*} > 0} \beta_{n} (\lambda - x_{n*}) \]
III. Negative Binomial Counts

\[
I(\mu,k \mid \{x_{n}\}) \propto \prod_{n=1}^{N} \frac{\Gamma(k + x_{n\star})}{\Gamma(x_{n\star} + 1) \Gamma(k)} \left(\frac{k}{\mu + k}\right)^{x_{n\star}} \left(\frac{\mu}{\mu + k}\right)^{k}
\]

\[
-L = \sum_{n} \left( -\log\left(\Gamma(k + x_{n\star})\right) + \log\left(\Gamma(k)\right) + \log\left(\Gamma(x_{n\star} + 1)\right) \right)
\]

\[
-\frac{dL}{d\mu} = \sum_{n} \left( 1 - \frac{x_{n\star}}{\mu} + \log(\mu + k) \right)
\]

\[
-\frac{dL}{dk} = \sum_{n} (\psi(k) - \psi(k + x_{n\star}) + \log(\mu + k))
\]

IV. Poisson Mixture

\[
i(c,q,\lambda \mid \{x_{ij}\}) \propto \prod_{n=1}^{N} \frac{[1 - cq^{3}]^{x_{n1}} [cq^{3}(1 - q^{2})]^{x_{n2}} [cq^{5}(1 - q^{5})]^{x_{n3}}}{\Gamma(x_{n1} + 1) \Gamma(x_{n2} + 1) \Gamma(x_{n3} + 1)} \lambda^{x_{n\star}} e^{-\lambda(1-cq^{6})}
\]

\[
-L = \sum_{n} \left[ -x_{n1} \lambda^{1-cq^{3}} - (x_{n2} + x_{n3}) \lambda^{c} - (3x_{n2} + 5x_{n3}) \lambda^{q} - x_{n2} \lambda^{1-q^{2}} \right]
\]

\[
-\frac{dL}{dc} = \sum_{n} \left[ \frac{x_{n1}q^{3}}{1-cq^{3}} - \frac{x_{n2} + x_{n3}}{c} - \lambda q^{10} \right]
\]

\[
-\frac{dL}{dq} = \sum_{n} \left[ \frac{3x_{n1}cq^{2}}{1-cq^{3}} - \frac{3x_{n2} + 5x_{n3}}{q} + \frac{2x_{n2}q}{1-q^{2}} + \frac{5x_{n3}q^{4}}{1-q^{5}} - 10\lambda q^{9} \right]
\]

\[
-\frac{dL}{d\lambda} = \sum_{n} \left[ 1 - cq^{10} - \frac{x_{n\star}}{\lambda} \right]
\]
V. ZIP Mixture

For $x_{n*} = 0$:

$$l(c, q, \lambda, \phi | \{x_n\}) \propto \prod_{n=1}^{N} \left( \phi + (1 - \phi) e^{-\lambda(1-cq^{10})} \right)$$

$$-L = - \log \left( \phi + (1 - \phi) e^{-\lambda(1-cq^{10})} \right)$$

$$\frac{-dL}{dc} = - \sum_{n} (1 - \phi) e^{-\lambda(1-cq^{10})} \lambda q^{10}$$

$$\frac{-dL}{dq} = - \sum_{n} (1 - \phi) \cdot 10 \lambda c q^{9} e^{-\lambda(1-cq^{10})}$$

$$\frac{-dL}{d\lambda} = \sum_{n} (1 - \phi) e^{-\lambda(1-cq^{10})} (1 - c q^{10})$$

$$\frac{-dL}{d\phi} = - \sum_{n} 1 - e^{-\lambda(1-cq^{10})}$$

For $x_{n*} > 0$:

$$l(c, q, \lambda, \phi | \{x_n\}) \propto \prod_{n=1}^{N} \left( \phi + (1 - \phi) \frac{[1-cq^{10}]^{x_{n1}} [cq^{3}(1-q^{10})]^{x_{n2}} [cq^{5}(1-q^{5})]^{x_{n3}}}{\Gamma(x_{n1}+1)\Gamma(x_{n2}+1)\Gamma(x_{n3}+1)} \lambda^{x_{n1}} e^{-\lambda(1-cq^{10})} \right)$$

$$-L = \sum_{n} \left[ - \log(1 - \phi) - x_{n} \ln(1-cq^{10}) - (3x_{2} + 5x_{3}) \ln(q) - (x_{2} + x_{3}) \ln(c) \right.$$

$$\left. - x_{2} \ln(1-q^{10}) - x_{3} \ln(1-q^{5}) - (x_{n*}) \ln(\lambda) + \lambda \left( 1-cq^{10} \right) + \sum_{j=1}^{3} \ln(\Gamma(x_{n}+1)) \right]$$

$$\frac{-dL}{dc} = \sum_{n} \left( \frac{x_{2}q^{3}}{1-cq^{10}} - \frac{x_{2} + x_{3}}{c} + \lambda q^{10} \right)$$

$$\frac{-dL}{dq} = \sum_{n} \left( \frac{3x_{2}q^{2}}{1-cq^{10}} - \frac{3x_{2} + 5x_{3}}{q} + \frac{2x_{2}q}{1-q^{10}} + \frac{5x_{2}q^{4}}{1-q^{5}} - 10\lambda c q^{9} \right)$$

$$\frac{-dL}{d\lambda} = \sum_{n} \left( 1-cq^{10} - \frac{x_{n*}}{\lambda} \right)$$

$$\frac{-dL}{d\phi} = \sum_{n} \frac{1}{1-\phi}$$
VI. NB Mixture

\[ l(c,q,k,\mu | \{ x_n \}) \propto \prod_{n=1}^{N} \left( \frac{1 - cq^3}{\Gamma(x_{n0} + 1)} \frac{1}{\Gamma(x_{n1} + 1)} \frac{1}{\Gamma(x_{n2} + 1)} \frac{1}{\Gamma(x_{n3} + 1)} \frac{\mu}{\mu + k} \right)^{x_{n0}} \left( \frac{k}{\mu + k} \right)^k \]

\[ L = \sum_{n} \left[ -x_{n1} \log(1 - cq^3) - (x_{n2} + x_{n3}) \log(c) - (3x_{n2} + 5x_{n3}) \log(q) - x_{n2} \log(1 - q^2) - x_{n3} \log(1 - q^2) + (x_{n0} + k) \log(\mu + k) - x_{n0} \log(\mu) - k \log(k) + \log(\Gamma(k)) \right] \]

As you might expect, this likelihood takes the longest to maximize (minimize). The time to convergence depends strongly on the upper limit of summation in the final term (which is infinite in the formula). In practice, this term decays fairly quickly so that even iterating out to \(x_{n0}=25\) is wasteful of cpu time if carried out every time the function is evaluated. Thus it speeds things up considerably to allow this loop to be exited early when the term has decayed past the point at which it can contribute any meaningful information. We set this limit to be the ratio of the summation term (evaluated for a particular value of \(x_{n0}\) summed over all data points) to the rest of the likelihood (summed over all data points and including the summation term evaluated up to \(x_{n0}-1\)). We stopped evaluating the likelihood when this ratio was less than a hundredth of the tolerance of the minimization algorithm (set to 10\(^{-8}\)).
\[- \frac{dL}{dk} = \sum_{n} \left[ \frac{x_{n^*} + k}{\mu + k} \log(\mu + k) - 1 - \log(k) + \psi(k) \right] \]
\[\times \left( \sum_{x_{n^*} = 0}^{\infty} \frac{\left( \mu c q^0 \right)^{x_{n^*}}}{\Gamma(x_{n^*} + 1)} \frac{\Gamma(x_{n^*} + x_{n^*} + k)}{(\mu + k)^{x_{n^*}} \Gamma(x_{n^*} + 1)} \left( \psi(x_{n^*} + x_{n^*} + k) - \frac{x_{n^*}}{\mu + k} \right) \right)^{-1} \]