1 Likelihood

- S_t: the t-th sample.
- $n_t := |S_t|$: sample size at each sample draw. Depending on the sampling scheme, it is possible for n_t to be random. For simplicity, assume that $n_t = n$ is fixed.
- $S_T^{in} = \bigcup_{t=1}^{T} S_t$: set of sampled individuals up to and including T-th sample.
- S_T^{out}: set of individuals that has not been sampled after T sample draws. ($S_T^{out} \cap S_T^{in} = \emptyset$).
- $\mathcal{P} = S_T^{in} \cup S_T^{out}$: the entire population.
- $d_{i(T)}$: number of samples that contained individual i:

$$d_{i(T)} = \sum_{t=1}^{T} \mathbb{1}\{i \in S_t\}$$

Assuming that samples are drawn independently:

$$\mathbb{E}[d_{i(T)}] = \sum_{t=1}^{T} \pi_i = T\pi_i$$

where π_i denotes inclusion probability of i. This suggests estimating π_i by

$$\hat{\pi}_i = \frac{d_{i(T)}}{T} \quad \text{or} \quad \hat{\pi}_i = \frac{1 + d_{i(T)}}{1 + T} \tag{1}$$

where the latter comes from enforcing the probabilities to be non-zero. Note that in this case $\hat{\pi}_i = \frac{1}{1+T} \forall i \in S_T^{out}$.

1.1 Homogenous inclusion probabilities

Let $\pi_i = \frac{n}{|\mathcal{P}|}\ \forall i = 1, \ldots, |\mathcal{P}|$. Then the inclusion frequencies follow a binomial distribution.

$$d_{i(T)} \overset{i.i.d.}{\sim} Bin(T, \frac{n}{|\mathcal{P}|})$$

Note that we do not observe $d_{i(T)} = 0$, but only $d_{i(T)} > 0$. Truncating the distribution at 0 yields

$$\begin{aligned}
\mathbb{P}(d_{i(T)} = k | d_{i(T)} > 0) = \frac{\mathbb{P}(d_{i(T)} = k)}{1 - \mathbb{P}(d_{i(T)} = 0)} &= \frac{\binom{T}{k}(\frac{n}{|\mathcal{P}|})^k(\frac{|\mathcal{P}|-n}{|\mathcal{P}|})^{T-k}}{1 - (\frac{|\mathcal{P}|-n}{|\mathcal{P}|})^T} \\
= \binom{T}{k}\frac{n^k(|\mathcal{P}| - n)^{T-k}}{|\mathcal{P}|^T - (|\mathcal{P}| - n)^T} &= \binom{T}{k}\frac{n^k(|S_T^{in}| + |S_T^{out}| - n)^{T-k}}{(|S_T^{in}| + |S_T^{out}|)^T - (|S_T^{in}| + |S_T^{out}| - n)^T}
\end{aligned}$$

This leads to the joint likelihood

$$\begin{aligned}
L(|S_T^{out}|) &:= L(|S_T^{out}|; T, d_{i(T)} = k_i \forall i \in S_T^{in}) \\
&= \prod_{i \in S_T^{in}} \binom{T}{k_i} \frac{n^{k_i}(|S_T^{in}| + |S_T^{out}| - n)^{T-k_i}}{(|S_T^{in}| + |S_T^{out}|)^T - (|S_T^{in}| + |S_T^{out}| - n)^T} \\
&\propto \prod_{i \in S_T^{in}} \frac{(|S_T^{in}| + |S_T^{out}| - n)^{T-k_i}}{(|S_T^{in}| + |S_T^{out}|)^T - (|S_T^{in}| + |S_T^{out}| - n)^T} \\
&= [(|S_T^{in}| + |S_T^{out}|)^T - (|S_T^{in}| + |S_T^{out}| - n)^T]^{-|S_T^{in}|}(|S_T^{in}| + |S_T^{out}| - n)^{T(|S_T^{in}|-n)}
\end{aligned}$$

since $\sum_{i \in S_T^{in}} k_i = nT$ for fixed n. Taking logarithm of the likelihood leads to

$$\begin{aligned} l(|S_T^{out}|) &:= \log L(|S_T^{out}|) \\ &= const - |S_T^{in}| \log[(|S_T^{in}| + |S_T^{out}|)^T - (|S_T^{in}| + |S_T^{out}| - n)^T] \\ &+ T(|S_T^{in}| - n) \log(|S_T^{in}| + |S_T^{out}| - n) \end{aligned} \tag{2}$$

The derivative with respect to $|S_T^{out}|$ is

$$\begin{aligned} \frac{d}{d|S_T^{out}|} l(|S_T^{out}|) = &- |S_T^{in}| T \frac{(|S_T^{in}| + |S_T^{out}|)^{T-1} - (|S_T^{in}| + |S_T^{out}| - n)^{T-1}}{(|S_T^{in}| + |S_T^{out}|)^T - (|S_T^{in}| + |S_T^{out}| - n)^T} \\ &+ \frac{T(|S_T^{in}| - n)}{|S_T^{out}| + |S_T^{in}| - n} \end{aligned} \tag{3}$$

1.1.1 Classical capture-recapture

Setting $T = 2$ results in the following log-likelihood:

$$\begin{aligned} l(|S_T^{out}|) = &\, const - |S_T^{in}| \log[(|S_T^{in}| + |S_T^{out}|)^2 - (|S_T^{in}| + |S_T^{out}| - n)^2] \\ &+ 2(|S_T^{in}| - n) \log(|S_T^{in}| + |S_T^{out}| - n) \end{aligned}$$

The first-order condition is

$$\frac{d}{d|S_T^{out}|} l(|S_T^{out}|) = - \frac{2|S_T^{in}|n}{(|S_T^{in}| + |S_T^{out}|)^2 - (|S_T^{in}| + |S_T^{out}| - n)^2} + \frac{2(|S_T^{in}| - n)}{|S_T^{out}| + |S_T^{in}| - n} \stackrel{!}{=} 0$$

Then, the maximum likelihood estimator of $|S_T^{out}|$ can be expressed as

$$\widehat{|S_T^{out}|}_{ML} = \frac{(n - |S_2^{in}|)(|S_2^{in}| - n)}{|S_2^{in}| - 2n} \tag{4}$$

1.1.2 One draw

If $T = 1$, then $|S_T^{in}| = \sum_{i \in S_T^{in}} k_i = n$. The log-likelihood becomes

$$l(|S_T^{out}|) = const - |S_T^{in}| \log(|S_T^{in}|) =: const$$

1.2 Heterogenous inclusion probabilities

Consider the problem from the Bayesian perspective. Assume Beta prior for inclusion probabilities:

$$\pi_i \sim Be(\alpha, \beta)$$

Then

$$\mathbb{E}[\sum_{i \in \mathcal{P}} \pi_i] = \sum_{i \in \mathcal{P}} \frac{\alpha}{\alpha + \beta} = |\mathcal{P}| \frac{\alpha}{\alpha + \beta} = (|S_T^{out}| + |S_T^{in}|) \frac{\alpha}{\alpha + \beta}$$

$$\mathbb{E}[\sum_{i \in \mathcal{P}} \pi_i] = \mathbb{E}[n_t] \Leftrightarrow (|S_T^{out}| + |S_T^{in}|) \frac{\alpha}{\alpha + \beta} = n \tag{5}$$

(in case of random n_t, we can estimate $\mathbb{E}[n_t]$ by $T^{-1} \sum_{t=1}^{T} n_t$).
Condition (5) will be the constraint for our model.
The likelihood would be

$$d_{i(T)} | \pi_i \sim Bin(T, \pi)$$

Then the posterior distribution is

$$\begin{aligned}
f(\pi_i|d_{i(T)} = k) &= \frac{\mathbb{P}(d_{i(T)} = k|\pi_i)f(\pi_i)}{\mathbb{P}(d_{i(T)} = k)} \\
&\propto \pi_i^k(1-\pi_i)^{T-k}\pi_i^{\alpha-1}(1-\pi_i)^{\beta-1} \\
&= \pi_i^{\alpha+k-1}(1-\pi_i)^{\beta+T-k-1} \\
&\Rightarrow \pi_i|d_{i(T)} = k \sim Be(\alpha+k, \beta+T-k) \\
&\Rightarrow \mathbb{E}[\pi_i|d_{i(T)} = k] = \frac{\alpha+k}{\alpha+\beta+T}
\end{aligned}$$

The marginal likelihood is:

$$\begin{aligned}
\mathbb{P}(d_{i(T)} = k) &= \int_0^1 f(\pi_i, d_{i(T)})d\pi_i = \int_0^1 \mathbb{P}(d_{i(T)} = k|\pi_i)f(\pi_i)d\pi_i \\
&= \binom{T}{k}\frac{\mathrm{B}(\alpha+k, \beta+T-k)}{\mathrm{B}(\alpha,\beta)}
\end{aligned}$$

Since we never observe $d_{i(T)} = 0$, the observed frequences $d_{i(T)} > 0$ for $i \in S_T^{in}$ follow a truncated distribution:

$$\begin{aligned}
\mathbb{P}(d_{i(T)} = k|d_{i(T)} > 0) &= \begin{cases} \frac{\mathbb{P}(d_{i(T)}=k)}{1-\mathbb{P}(d_{i(T)}=0)}, & \text{if } k > 0 \\ 0, & \text{otherwise} \end{cases} \\
&= \begin{cases} \binom{T}{k}\frac{\mathrm{B}(\alpha+k,\beta+T-k)}{\mathrm{B}(\alpha,\beta)-\mathrm{B}(\alpha,\beta+T)}, & \text{if } k > 0 \\ 0, & \text{otherwise} \end{cases}
\end{aligned}$$

Following empirical Bayes approach, maximise the marginal likelihood to obtain hyper-parameters α and β.

$$\begin{aligned}
L(\alpha,\beta) &:= L(\alpha,\beta; T, d_{i(T)} = k_i \forall i \in S_T^{in}) \\
&\overset{indep}{=} \prod_{i\in S_T^{in}} \binom{T}{k_i}\frac{\Gamma(\alpha+k_i)\Gamma(\beta+T-k_i)}{\Gamma(\alpha+\beta+T)} \cdot \frac{1}{\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} - \frac{\Gamma(\alpha)\Gamma(\beta+T)}{\Gamma(\alpha+\beta+T)}} \\
&= \prod_{i\in S_T^{in}} \binom{T}{k_i}\frac{\Gamma(\alpha+k_i)\Gamma(\beta+T-k_i)\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha+\beta+T) - \Gamma(\alpha)\Gamma(\beta+T)\Gamma(\alpha+\beta)}
\end{aligned} \tag{6}$$

Using the recursive formula of gamma function $\Gamma(x) = (x-1)\Gamma(x-1)$ and the fact that $T, k_i \in \mathbb{N}_0 := \mathbb{N} \cup \{0\}$ allows us to rewrite the marginal likelihood as:

$$\begin{aligned}
L(\alpha,\beta) &\propto \prod_{i\in S_T^{in}} \frac{\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha+\beta)} \\
&\quad \times \frac{\prod_{j=1}^{k_i}(\alpha+k_i-j)\prod_{j=1}^{T-k_i}(\beta+T-k_i-j)}{\prod_{j=1}^{T}(\alpha+\beta+T-j) - \prod_{j=1}^{T}(\beta+T-j)} \\
&= \prod_{i\in S_T^{in}} \frac{\prod_{j=1}^{k_i}(\alpha+k_i-j)\prod_{j=1}^{T-k_i}(\beta+T-k_i-j)}{\prod_{j=1}^{T}(\alpha+\beta+T-j) - \prod_{j=1}^{T}(\beta+T-j)}
\end{aligned}$$

The parameters α and β must satisfy (5), so equation (6) is to be maximised subject to the constraint. Note that (5) can be viewed as enforcing our prior to be centered at $\frac{n}{|\mathcal{P}|}$, since

$$(|S_T^{out}| + |S_T^{in}|)\frac{\alpha}{\alpha+\beta} = n \Leftrightarrow \frac{\alpha}{\alpha+\beta} = \mathbb{E}[\pi_i] = \frac{n}{(|S_T^{out}| + |S_T^{in}|)}$$

Intuitively, this means that, under the constraint, the hyperparameters will determine only the variance of the prior.

The optimisation problem can be simplified by solving the constraint for β and plugging into the objective function:

$$\begin{aligned} &\max_{\alpha,\beta} L(\alpha,\beta) \quad \text{s.t.} \quad |\mathcal{P}|\frac{\alpha}{\alpha+\beta} = n \\ &\qquad\qquad \Leftrightarrow \beta = \left(\frac{|\mathcal{P}|}{n} - 1\right)\alpha := q\alpha \\ \Rightarrow &\max_{\alpha,q} L(\alpha,q) \end{aligned}$$

More explicitly,

$$\begin{aligned} &\max_{\alpha,\beta} L(\alpha,\beta) \quad \text{s.t.} \quad (|S_T^{in}| + |S_T^{out}|)\frac{\alpha}{\alpha+\beta} = n \\ &\qquad\qquad \Leftrightarrow \beta = n^{-1}\alpha|S_T^{in}| + n^{-1}\alpha|S_T^{out}| - \alpha \\ \Rightarrow &\max_{\alpha,|S_T^{out}|} L(\alpha,|S_T^{out}|) \end{aligned}$$

1.2.1 Variant 1

$$L(\alpha,q) \propto \prod_{i\in S_T^{in}} \frac{\prod_{j=1}^{k_i}(\alpha+k_i-j)\prod_{j=1}^{T-k_i}(q\alpha+T-k_i-j)}{\prod_{j=1}^{T}(\alpha+q\alpha+T-j) - \prod_{j=1}^{T}(q\alpha+T-j)}$$

$$\begin{aligned} l(\alpha,q) &:= \log L(\alpha,q) \\ &= const - |S_T^{in}| \log\left[\prod_{j=1}^{T}(\alpha+q\alpha+T-j) - \prod_{j=1}^{T}(q\alpha+T-j)\right] \\ &+ \sum_{i\in S_T^{in}}\sum_{j=1}^{k_i}\log(\alpha+k_i-j) + \sum_{j=1}^{T-k_i}\log(q\alpha+T-k_i-j) \end{aligned} \tag{7}$$

Derivatives:

$$
\begin{aligned}
\frac{\partial l(\alpha,q)}{\partial \alpha} =& -|S_T^{in}|\frac{\sum_{j=1}^T (1+q)\prod_{1\le m\neq j\le T}(\alpha+q\alpha+T-m)-\sum_{j=1}^T q\prod_{1\le m\neq j\le T}(q\alpha+T-m)}{\prod_{j=1}^T(\alpha+q\alpha+T-j)-\prod_{j=1}^T(q\alpha+T-j)} \\
&+\sum_{i\in S_T^{in}}\sum_{j=1}^{k_i}(\alpha+k_i-j)^{-1}+q\sum_{j=1}^{T-k_i}(q\alpha+T-k_i-j)^{-1} \\
=& -(1+q)|S_T^{in}|\frac{\prod_{j=1}^T(\alpha+q\alpha+T-j)\sum_{j=1}^T(\alpha+q\alpha+T-j)^{-1}}{\prod_{j=1}^T(\alpha+q\alpha+T-j)-\prod_{j=1}^T(q\alpha+T-j)} \\
&+q|S_T^{in}|\frac{\prod_{j=1}^T(q\alpha+T-j)\sum_{j=1}^T(q\alpha+T-j)^{-1}}{\prod_{j=1}^T(\alpha+q\alpha+T-j)-\prod_{j=1}^T(q\alpha+T-j)} \\
&+\sum_{i\in S_T^{in}}\sum_{j=1}^{k_i}(\alpha+k_i-j)^{-1}+q\sum_{j=1}^{T-k_i}(q\alpha+T-k_i-j)^{-1} \\
=& -(1+q)|S_T^{in}|\frac{A_p(\alpha,q)A_s(\alpha,q)}{A_p(\alpha,q)-B_p(\alpha,q)}+q|S_T^{in}|\frac{B_p(\alpha,q)B_s(\alpha,q)}{A_p(\alpha,q)-B_p(\alpha,q)} \\
&+\sum_{i\in S_T^{in}}\sum_{j=1}^{k_i}(\alpha+k_i-j)^{-1}+q\sum_{j=1}^{T-k_i}(q\alpha+T-k_i-j)^{-1}
\end{aligned}
\tag{8}
$$

$$
\begin{aligned}
\frac{\partial l(\alpha,q)}{\partial q} =& -|S_T^{in}|\frac{\sum_{j=1}^T \alpha\prod_{1\le m\neq j\le T}(\alpha+q\alpha+T-m)-\sum_{j=1}^T \alpha\prod_{1\le m\neq j\le T}(q\alpha+T-m)}{\prod_{j=1}^T(\alpha+q\alpha+T-j)-\prod_{j=1}^T(q\alpha+T-j)} \\
&+\sum_{i\in S_T^{in}}\alpha\sum_{j=1}^{T-k_i}(q\alpha+T-k_i-j)^{-1} \\
=& -\alpha|S_T^{in}|\frac{\prod_{j=1}^T(\alpha+q\alpha+T-j)\sum_{j=1}^T(\alpha+q\alpha+T-j)^{-1}}{\prod_{j=1}^T(\alpha+q\alpha+T-j)-\prod_{j=1}^T(q\alpha+T-j)} \\
&+\alpha|S_T^{in}|\frac{\prod_{j=1}^T(q\alpha+T-j)\sum_{j=1}^T(q\alpha+T-j)^{-1}}{\prod_{j=1}^T(\alpha+q\alpha+T-j)-\prod_{j=1}^T(q\alpha+T-j)} \\
&+\alpha\sum_{i\in S_T^{in}}\sum_{j=1}^{T-k_i}(q\alpha+T-k_i-j)^{-1} \\
=& -\alpha|S_T^{in}|\frac{A_p(\alpha,q)A_s(\alpha,q)-B_p(\alpha,q)B_s(\alpha,q)}{A_p(\alpha,q)-B_p(\alpha,q)}+\alpha\sum_{i\in S_T^{in}}\sum_{j=1}^{T-k_i}(q\alpha+T-k_i-j)^{-1}
\end{aligned}
\tag{9}
$$

where

$$
A_p(\alpha,q):=\prod_{j=1}^T(\alpha+q\alpha+T-j),\quad A_s(\alpha,q):=\sum_{j=1}^T(\alpha+q\alpha+T-j)^{-1}
$$

$$
B_p(\alpha,q):=\prod_{j=1}^T(q\alpha+T-j),\quad B_s(\alpha,q):=\sum_{j=1}^T(q\alpha+T-j)^{-1}
$$

The product terms $A_p(\alpha, q)$ and $B_p(\alpha, q)$ are computationally expensive to calculate, as even small values of T will lead yield extremely large quantities. For practical purposes, one can factor out $q\alpha + T$ from both expressions. Then the log-likelihood and the derivatives can be re-written as

$$\begin{aligned} l(\alpha, q) &= const - T|S_T^{in}| \log(q\alpha + T) - |S_T^{in}| \log\left[\widetilde{A_p}(\alpha, q) - \widetilde{B_p}(\alpha, q)\right] \\ &+ \sum_{i \in S_T^{in}} \sum_{j=1}^{k_i} \log(\alpha + k_i - j) + \sum_{j=1}^{T-k_i} \log(q\alpha + T - k_i - j) \end{aligned}$$

$$\begin{aligned} \frac{\partial l(\alpha, q)}{\partial \alpha} =& -(1+q)|S_T^{in}| \frac{\widetilde{A_p}(\alpha, q) A_s(\alpha, q)}{\widetilde{A_p}(\alpha, q) - \widetilde{B_p}(\alpha, q)} + q|S_T^{in}| \frac{\widetilde{B_p}(\alpha, q) B_s(\alpha, q)}{\widetilde{A_p}(\alpha, q) - \widetilde{B_p}(\alpha, q)} \\ &+ \sum_{i \in S_T^{in}} \sum_{j=1}^{k_i} (\alpha + k_i - j)^{-1} + q \sum_{j=1}^{T-k_i} (q\alpha + T - k_i - j)^{-1} \end{aligned}$$

$$\frac{\partial l(\alpha, q)}{\partial q} = -\alpha|S_T^{in}| \frac{\widetilde{A_p}(\alpha, q) A_s(\alpha, q) - \widetilde{B_p}(\alpha, q) B_s(\alpha, q)}{\widetilde{A_p}(\alpha, q) - \widetilde{B_p}(\alpha, q)} + \alpha \sum_{i \in S_T^{in}} \sum_{j=1}^{T-k_i} (q\alpha + T - k_i - j)^{-1}$$

where

$$\widetilde{A_p}(\alpha, q) := \prod_{j=1}^{T} \left(1 - \frac{j - \alpha}{q\alpha + T}\right), \quad \widetilde{B_p}(\alpha, q) := \prod_{j=1}^{T} \left(1 - \frac{j}{q\alpha + T}\right)$$

1.2.2 Variant 2

$$L(\alpha, |S_T^{out}|) \propto \prod_{i \in S_T^{in}} \frac{\prod_{j=1}^{k_i} (\alpha + k_i - j) \prod_{j=1}^{T-k_i} (n^{-1}\alpha|S_T^{in}| + n^{-1}\alpha|S_T^{out}| - \alpha + T - k_i - j)}{\prod_{j=1}^{T} (n^{-1}\alpha|S_T^{in}| + n^{-1}\alpha|S_T^{out}| + T - j) - \prod_{j=1}^{T} (n^{-1}\alpha|S_T^{in}| + n^{-1}\alpha|S_T^{out}| - \alpha + T - j)}$$

$$\begin{aligned} l(\alpha, |S_T^{out}|) &:= \log L(\alpha, |S_T^{out}|) \\ &= const - |S_T^{in}| \log\left[A_p(\alpha, |S_T^{out}|) - B_p(\alpha, |S_T^{out}|)\right] \\ &+ \sum_{i \in S_T^{in}} \sum_{j=1}^{k_i} \log(\alpha + k_i - j) + \sum_{j=1}^{T-k_i} \log(n^{-1}\alpha|S_T^{in}| + n^{-1}\alpha|S_T^{out}| - \alpha + T - k_i - j) \\ &= const - T|S_T^{in}| \log(n^{-1}\alpha|S_T^{in}| + n^{-1}\alpha|S_T^{out}| + T) - |S_T^{in}| \log\left[\widetilde{A_p}(\alpha, |S_T^{out}|) - \widetilde{B_p}(\alpha, |S_T^{out}|)\right] \\ &+ \sum_{i \in S_T^{in}} \sum_{j=1}^{k_i} \log(\alpha + k_i - j) + \sum_{j=1}^{T-k_i} \log(n^{-1}\alpha|S_T^{in}| + n^{-1}\alpha|S_T^{out}| - \alpha + T - k_i - j) \end{aligned} \tag{10}$$

Derivatives:

$$
\begin{aligned}
\frac{\partial l(\alpha,|S_T^{out}|)}{\partial \alpha} =& -|S_T^{in}|(n^{-1}|S_T^{in}|+n^{-1}|S_T^{out}|)\frac{A_p(\alpha,|S_T^{out}|)A_s(\alpha,|S_T^{out}|)}{A_p(\alpha,|S_T^{out}|)-B_p(\alpha,|S_T^{out}|)}\\
&+|S_T^{in}|(n^{-1}|S_T^{in}|+n^{-1}|S_T^{out}|-1)\frac{B_p(\alpha,|S_T^{out}|)B_s(\alpha,|S_T^{out}|)}{A_p(\alpha,|S_T^{out}|)-B_p(\alpha,|S_T^{out}|)}\\
&+\sum_{i\in S_T^{in}}\sum_{j=1}^{k_i}(\alpha+k_i-j)^{-1}+(\frac{|S_T^{in}|}{n}+\frac{|S_T^{out}|}{n}-1)\sum_{j=1}^{T-k_i}(\alpha\frac{|S_T^{in}|}{n}+\alpha\frac{|S_T^{out}|}{n}-\alpha+T-k_i-j)^{-1}\\
=& -|S_T^{in}|(n^{-1}|S_T^{in}|+n^{-1}|S_T^{out}|)\frac{\widetilde{A_p}(\alpha,|S_T^{out}|)A_s(\alpha,|S_T^{out}|)}{\widetilde{A_p}(\alpha,|S_T^{out}|)-\widetilde{B_p}(\alpha,|S_T^{out}|)}\\
&+|S_T^{in}|(n^{-1}|S_T^{in}|+n^{-1}|S_T^{out}|-1)\frac{\widetilde{B_p}(\alpha,|S_T^{out}|)B_s(\alpha,|S_T^{out}|)}{\widetilde{A_p}(\alpha,|S_T^{out}|)-\widetilde{B_p}(\alpha,|S_T^{out}|)}\\
&+\sum_{i\in S_T^{in}}\sum_{j=1}^{k_i}(\alpha+k_i-j)^{-1}+(\frac{|S_T^{in}|}{n}+\frac{|S_T^{out}|}{n}-1)\sum_{j=1}^{T-k_i}(\alpha\frac{|S_T^{in}|}{n}+\alpha\frac{|S_T^{out}|}{n}-\alpha+T-k_i-j)^{-1}
\end{aligned}
\tag{11}
$$

$$
\begin{aligned}
\frac{\partial l(\alpha,|S_T^{out}|)}{\partial |S_T^{out}|} =& -n^{-1}\alpha|S_T^{in}|\frac{A_p(\alpha,|S_T^{out}|)A_s(\alpha,|S_T^{out}|)-B_p(\alpha,|S_T^{out}|)B_s(\alpha,|S_T^{out}|)}{A_p(\alpha,|S_T^{out}|)-B_p(\alpha,|S_T^{out}|)}\\
&+n^{-1}\alpha\sum_{i\in S_T^{in}}\sum_{j=1}^{T-k_i}(\alpha\frac{|S_T^{in}|}{n}+\alpha\frac{|S_T^{out}|}{n}+T-k_i-j)^{-1}\\
=& -n^{-1}\alpha|S_T^{in}|\frac{\widetilde{A_p}(\alpha,|S_T^{out}|)A_s(\alpha,|S_T^{out}|)-\widetilde{B_p}(\alpha,|S_T^{out}|)B_s(\alpha,|S_T^{out}|)}{\widetilde{A_p}(\alpha,|S_T^{out}|)-\widetilde{B_p}(\alpha,|S_T^{out}|)}\\
&+n^{-1}\alpha\sum_{i\in S_T^{in}}\sum_{j=1}^{T-k_i}(\alpha\frac{|S_T^{in}|}{n}+\alpha\frac{|S_T^{out}|}{n}+T-k_i-j)^{-1}
\end{aligned}
\tag{12}
$$

where

$$
\begin{aligned}
A_p(\alpha,|S_T^{out}|) &:= \prod_{j=1}^{T}(n^{-1}\alpha|S_T^{in}|+n^{-1}\alpha|S_T^{out}|+T-j)\\
A_s(\alpha,|S_T^{out}|) &:= \sum_{j=1}^{T}(n^{-1}\alpha|S_T^{in}|+n^{-1}\alpha|S_T^{out}|+T-j)^{-1}\\
B_p(\alpha,|S_T^{out}|) &:= \prod_{j=1}^{T}(n^{-1}\alpha|S_T^{in}|+n^{-1}\alpha|S_T^{out}|-\alpha+T-j)\\
B_s(\alpha,|S_T^{out}|) &:= \sum_{j=1}^{T}(n^{-1}\alpha|S_T^{in}|+n^{-1}\alpha|S_T^{out}|-\alpha+T-j)^{-1}
\end{aligned}
$$

and

$$\widetilde{A_p}(\alpha, |S_T^{out}|) := \prod_{j=1}^{T}\left(1 - \frac{j}{n^{-1}\alpha|S_T^{in}| + n^{-1}\alpha|S_T^{out}| + T}\right)$$
$$\widetilde{B_p}(\alpha, |S_T^{out}|) := \prod_{j=1}^{T}\left(1 - \frac{j+\alpha}{n^{-1}\alpha|S_T^{in}| + n^{-1}\alpha|S_T^{out}| + T}\right)$$

1.3 Capture-recapture with $T = 2$ and $\alpha \to \infty$

Consider the likelihood function from 1.2.2 with $T = 2$.

$$L(\alpha, |S_2^{in}|) \propto \prod_{i\in S_2^{in}} \frac{\prod_{j=1}^{k_i}(\alpha + k_i - j)\prod_{j=1}^{2-k_i}(n^{-1}\alpha|S_2^{in}| + n^{-1}\alpha|S_2^{out}| - \alpha + 2 - k_i - j)}{\prod_{j=1}^{2}(n^{-1}\alpha|S_2^{in}| + n^{-1}\alpha|S_2^{out}| + 2 - j) - \prod_{j=1}^{2}(n^{-1}\alpha|S_2^{in}| + n^{-1}\alpha|S_2^{out}| - \alpha + 2 - j)}$$

Since at $T = 2$ observed frequencies are $k_i \in \{1, 2\}$ for all $i \in S_2^{in}$, let us partition S_2^{in} into two disjoint sets $R := \{i \in S_2^{in} : k_i = 2\}$ (recaptured) and $C := \{i \in S_2^{in} : k_i = 1\}$ (captured). Then, the likelihood can be re-written as

$$L(\alpha, |S_2^{out}|) \propto$$
$$\prod_{i\in C} \frac{\alpha(n^{-1}|S_2^{in}| + n^{-1}|S_2^{out}| - 1)}{(\frac{\alpha}{n}|S_2^{in}| + \frac{\alpha}{n}|S_2^{out}| + 1)(\frac{1}{n}|S_2^{in}| + \frac{1}{n}|S_2^{out}|) - (\frac{\alpha}{n}|S_2^{in}| + \frac{\alpha}{n}|S_2^{out}| - \alpha + 1)(\frac{1}{n}|S_2^{in}| + \frac{1}{n}|S_2^{out}| - 1)}$$
$$\prod_{i\in R} \frac{\alpha + 1}{(\frac{\alpha}{n}|S_2^{in}| + \frac{\alpha}{n}|S_2^{out}| + 1)(\frac{1}{n}|S_2^{in}| + \frac{1}{n}|S_2^{out}|) - (\frac{\alpha}{n}|S_2^{in}| + \frac{\alpha}{n}|S_2^{out}| - \alpha + 1)(\frac{1}{n}|S_2^{in}| + \frac{1}{n}|S_2^{out}| - 1)}$$

In both terms, the denominator can be simplified to

$$(\frac{\alpha}{n}|S_2^{in}| + \frac{\alpha}{n}|S_2^{out}| + 1)(\frac{1}{n}|S_2^{in}| + \frac{1}{n}|S_2^{out}|) - (\frac{\alpha}{n}|S_2^{in}| + \frac{\alpha}{n}|S_2^{out}| - \alpha + 1)(\frac{1}{n}|S_2^{in}| + \frac{1}{n}|S_2^{out}| - 1)$$
$$= \frac{2\alpha}{n}|S_2^{in}| + \frac{2\alpha}{n}|S_2^{out}| - \alpha + 1 = \alpha(\frac{2}{n}|S_2^{in}| + \frac{2}{n}|S_2^{out}| - 1 + \frac{1}{\alpha})$$

So the likelihood can expressed as

$$L(\alpha, |S_2^{out}|) \propto \frac{\alpha^{|C|}(n^{-1}|S_2^{in}| + n^{-1}|S_2^{out}| - 1)^{|C|}(\alpha+1)^{|R|}}{\alpha^{|S_2^{in}|}(\frac{2}{n}|S_2^{in}| + \frac{2}{n}|S_2^{out}| - 1 + \frac{1}{\alpha})^{|S_2^{in}|}}$$
$$= \frac{P^{|C|}(\alpha+1)^{|R|}}{\alpha^{|R|}(2P + 1 + \frac{1}{\alpha})^{|S_2^{in}|}}$$

where $P = n^{-1}|S_2^{in}| + n^{-1}|S_2^{out}| - 1$. Take the limit for $\alpha \to \infty$:

$$\lim_{\alpha\to\infty} \frac{P^{|C|}(\alpha+1)^{|R|}}{\alpha^{|R|}(2P + 1 + \frac{1}{\alpha})^{|S_2^{in}|}} \overset{\text{L'Hôpital}}{=} \lim_{\alpha\to\infty} \frac{P^{|C|}|R|(\alpha+1)^{|R|-1}}{|R|\alpha^{|R|-1}(2P + 1 + \frac{1}{\alpha})^{|S_2^{in}|} - \alpha^{|R|-2}|S_2^{in}|(2P + 1 + \frac{1}{\alpha})^{|S_2^{in}|-1}}$$

2 Simulation results

We simulate capture-recapture procedure to produce $d_{i(T)}$ as follows:

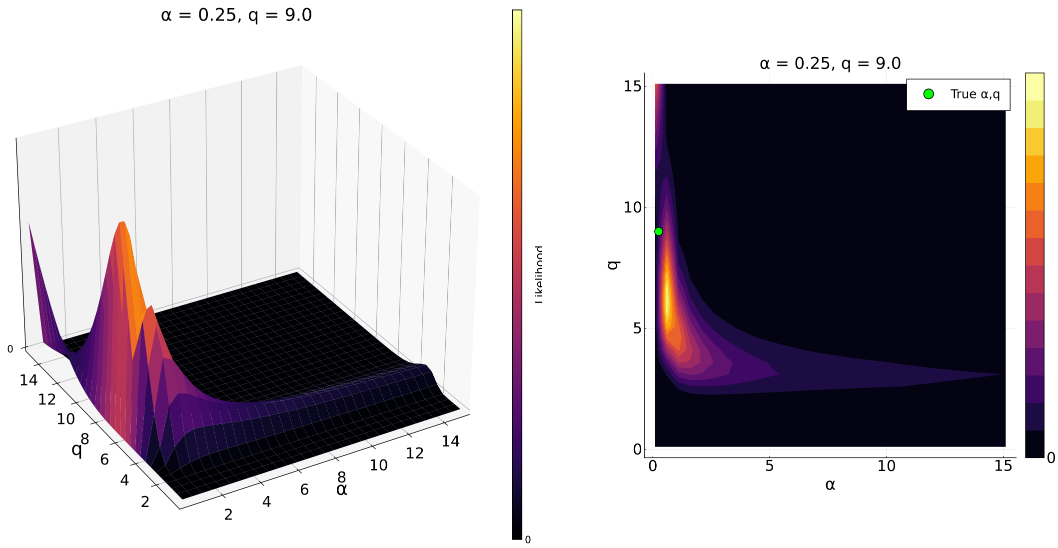

Figure 1: Likelihood function for data simulated by $\pi \sim Be(0.25, 9 \cdot 0.25), |\mathcal{P}| = 100, n = 10, T = 5$

1. Set the true population size to $|\mathcal{P}|$, sample size at each capture to n and number of captures to T. This results in setting $q = (\frac{\mathcal{P}}{n} - 1)$.
2. Set the parameter α.
3. Draw $|\mathcal{P}|$ values from $\pi \sim Beta(\alpha, q\alpha)$.
4. Normalise π_i by dividing by n.
5. At each replication $t = 1, \ldots, T$:
   (a) Given $\pi_i \forall\ i = 1, \ldots, |\mathcal{P}|$, draw a sample of n numbers from $\{1, \ldots, |\mathcal{P}|\}$ using Sampford's method. Denote the sample with S_t.
   (b) For each $i \in S_t$, increment $d_{i(T)}$ by 1. If $d_{i(T)}$ was never recorded before set $d_{i(T)} = 1$.

Once we acquire $d_{i(T)}$, it is possible to calculate the likelihood function from section 1.1. Figures 1-3 show surface and contour plots of the likelihood for data simulated using different true α at parameter values $\{0.1, 0.6, \ldots, 15.1\}$. The population size was set to 100, 5 samples of size 10 were drawn.

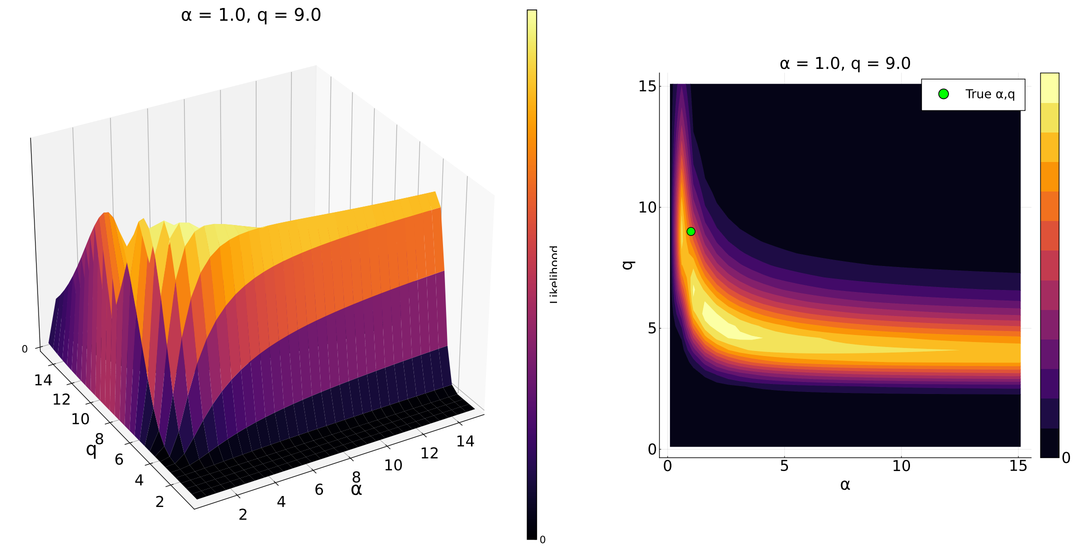

Figure 2: Likelihood function for data simulated by $\pi \sim Be(1, 9 \cdot 1), |\mathcal{P}| = 100, n = 10, T = 5$

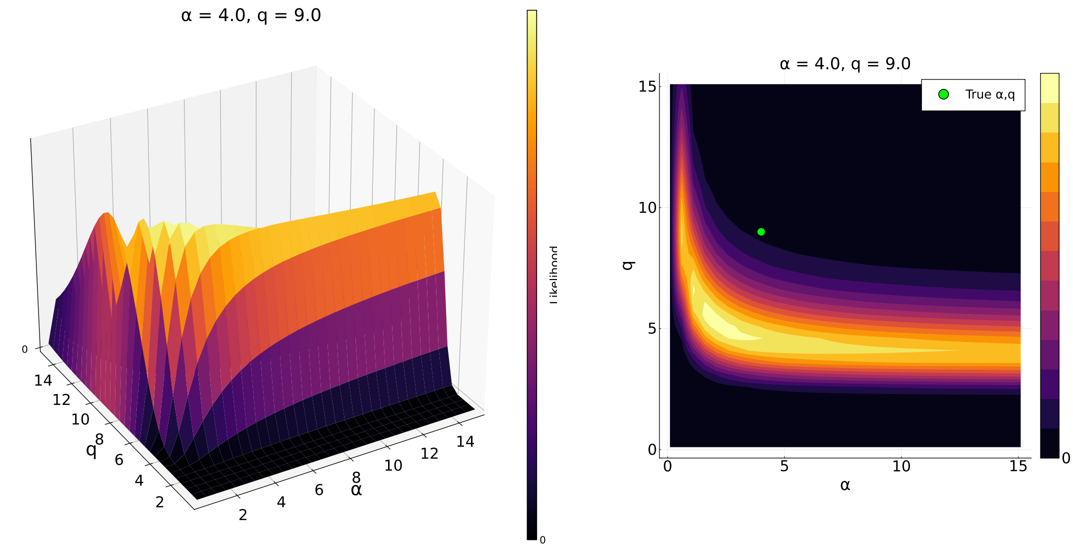

Figure 3: Likelihood function for data simulated by $\pi \sim Be(4, 9 \cdot 4), |\mathcal{P}| = 100, n = 10, T = 5$