1 Likelihood

- S_t: the t-th sample.
- $S_T^{in} = \bigcup_{t=1}^{T} S_t$: 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 $\hat{\pi}_i = \frac{1}{1+T} \forall i \in S_T^{out}$.

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

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

Then

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

Let $n_t := |S_t|$ be the sample size. While we assume independent replications of the same sampling scheme, depending on the chosen scheme, it is possible for n_t to be random. However, for simplicity, assume here that $n_t = n$ is fixed.

$$\mathbb{E}[\sum_{i \in \mathcal{P}} \pi_i] = \mathbb{E}[n_t] \Leftrightarrow (|S_T^{out}| + |S_T^{in}|) \frac{\alpha}{\alpha + \beta} = n \tag{2}$$

(in case of random n_t, we can estimate $\mathbb{E}[n_t]$ by $T^{-1} \sum_{t=1}^{T} n_t$).

Condition (2) 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:

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

Note that we never observe $d_{i(T)} = 0$. The observed frequences $d_{i(T)} > 0$ for $i \in S_T^{in}$ follow a truncated distribution:

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

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

$$\begin{aligned}L(\alpha, \beta) &:= L(\alpha, \beta; T, d_{i(T)} = k_i \forall i \in S_T^{in}) \\ &\overset{indep}{=} \prod_{i \in S_T^{in}} \binom{T}{k_i} \frac{\Gamma(\alpha + k_i)\Gamma(\beta + T - k_i)}{\Gamma(\alpha + \beta + T)} \cdot \frac{1}{\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} - \frac{\Gamma(\alpha)\Gamma(\beta+T)}{\Gamma(\alpha+\beta+T)}} \\ &= \prod_{i \in S_T^{in}} \binom{T}{k_i} \frac{\Gamma(\alpha + k_i)\Gamma(\beta + T - k_i)\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha + \beta + T) - \Gamma(\alpha)\Gamma(\beta + T)\Gamma(\alpha + \beta)}\end{aligned} \tag{3}$$

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:

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

The parameters α and β must satisfy (2), so equation (3) is to be maximised subject to the constraint. The optimisation problem can be simplified by solving the constraint for β and plugging into the objective function:

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

Alternatively,

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

1.1 Variant 1

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

$$\begin{aligned}l(\alpha,q) &:= \log L(\alpha,q) \\ &= const - |S_T^{in}|\log\left[\prod_{j=1}^{T}(\alpha+q\alpha+T-j) - \prod_{j=1}^{T}(q\alpha+T-j)\right] \\ &+ \sum_{i\in S_T^{in}}\sum_{j=1}^{k_i}\log(\alpha+k_i-j) + \sum_{j=1}^{T-k_i}\log(q\alpha+T-k_i-j)\end{aligned} \tag{4}$$

Derivatives:

$$\begin{aligned}\frac{\partial l(\alpha,q)}{\partial\alpha} =& -|S_{in}^T|\frac{\sum_{j=1}^{T}(1+q)\prod_{1\le m\neq j\le T}(\alpha+q\alpha+T-m) - \sum_{j=1}^{T}q\prod_{1\le m\neq j\le T}(q\alpha+T-m)}{\prod_{j=1}^{T}(\alpha+q\alpha+T-j)-\prod_{j=1}^{T}(q\alpha+T-j)} \\ &+ \sum_{i\in S_T^{in}}\sum_{j=1}^{k_i}(\alpha+k_i-j)^{-1} + q\sum_{j=1}^{T-k_i}(q\alpha+T-k_i-j)^{-1} \\ =& -(1+q)|S_{in}^T|\frac{\prod_{j=1}^{T}(\alpha+q\alpha+T-j)\sum_{j=1}^{T}(\alpha+q\alpha+T-j)^{-1}}{\prod_{j=1}^{T}(\alpha+q\alpha+T-j)-\prod_{j=1}^{T}(q\alpha+T-j)} \\ &+ q|S_T^{in}|\frac{\prod_{j=1}^{T}(q\alpha+T-j)\sum_{j=1}^{T}(q\alpha+T-j)^{-1}}{\prod_{j=1}^{T}(\alpha+q\alpha+T-j)-\prod_{j=1}^{T}(q\alpha+T-j)} \\ &+ \sum_{i\in S_T^{in}}\sum_{j=1}^{k_i}(\alpha+k_i-j)^{-1} + q\sum_{j=1}^{T-k_i}(q\alpha+T-k_i-j)^{-1}\end{aligned} \tag{5}$$

$$
\begin{aligned}
\frac{\partial l(\alpha,q)}{\partial \alpha} =& -|S_{in}^T|\frac{\sum_{j=1}^T \alpha \prod_{1\le m\neq j\le T}(\alpha+q\alpha+T-m)-\sum_{j=1}^T \alpha\prod_{1\le m\neq j\le T}(q\alpha+T-m)}{\prod_{j=1}^T(\alpha+q\alpha+T-j)-\prod_{j=1}^T(q\alpha+T-j)}\\
&+\sum_{i\in S_T^{in}}\alpha\sum_{j=1}^{T-k_i}(q\alpha+T-k_i-j)^{-1}\\
=& -\alpha|S_{in}^T|\frac{\prod_{j=1}^T(\alpha+q\alpha+T-j)\sum_{j=1}^T(\alpha+q\alpha+T-j)^{-1}}{\prod_{j=1}^T(\alpha+q\alpha+T-j)-\prod_{j=1}^T(q\alpha+T-j)}\\
&+\alpha|S_T^{in}|\frac{\prod_{j=1}^T(q\alpha+T-j)\sum_{j=1}^T(q\alpha+T-j)^{-1}}{\prod_{j=1}^T(\alpha+q\alpha+T-j)-\prod_{j=1}^T(q\alpha+T-j)}\\
&+\alpha\sum_{i\in S_T^{in}}\sum_{j=1}^{T-k_i}(q\alpha+T-k_i-j)^{-1}
\end{aligned}
\tag{6}
$$

1.2 Variant 2

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

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

Derivatives:

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

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

2 Simulation results

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

1. Set the true population size to N, sample size at each capture to n and number of captures to T.
2. Set the true parameters α and β.
3. Draw N values from $\pi \sim Beta(\alpha,\beta)$.
4. Normalise π_i by dividing by the sum $\sum_{i=1}^N \pi_i$.
5. At each replication $t = 1,\dots,T$:
   (a) Draw a sample of n numbers from $\{1,\dots,N\}$. Denote 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 exact likelihood using (3) multiplied by the prduct of binomial coefficients $\binom{T}{k_i}$. Figures 1-3 show surface and contour plots of the likelihood function for data simulated using different true α,β at parameter values $\{0.1, 0.6, \dots, 15.1\}$. The population size was set to 100, two samples of size 25 were drawn.

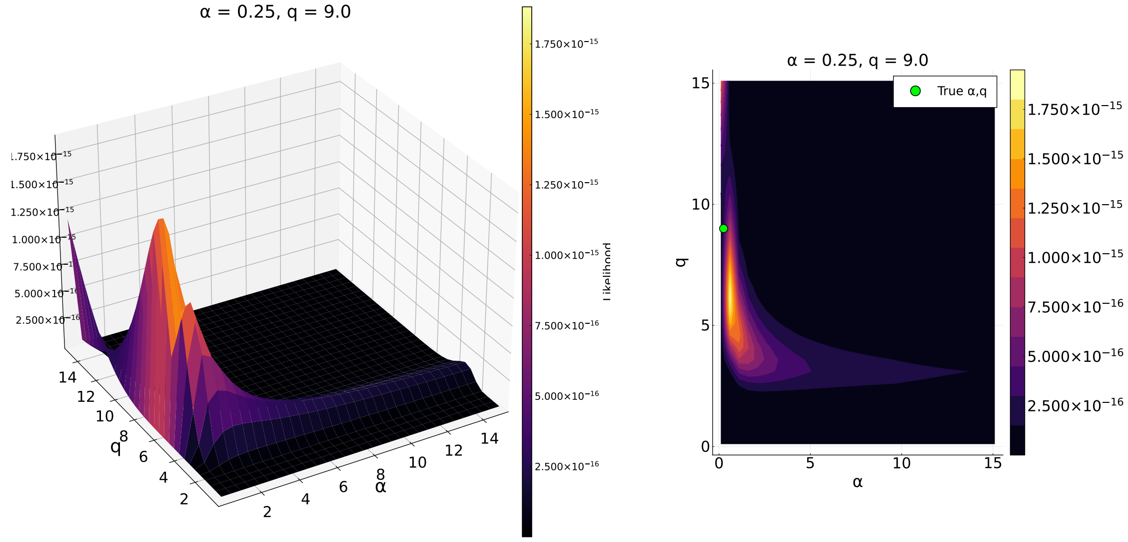

Figure 1: Likelihood function for data simulated by $\pi \sim Beta(1,4), N = 100, n = 25, T = 2$

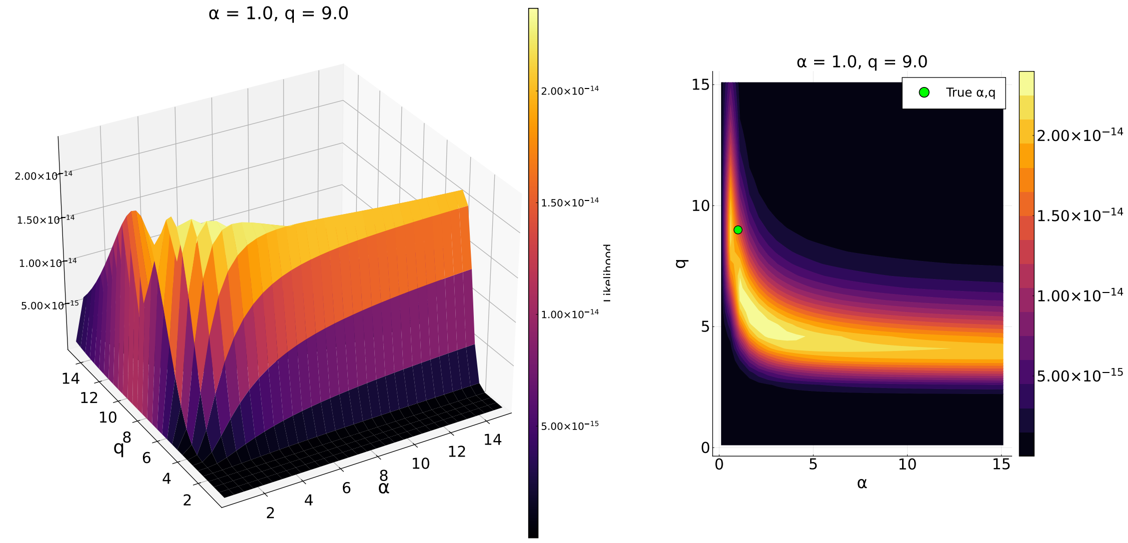

Figure 2: Likelihood function for data simulated by $\pi \sim Beta(1,1), N = 100, n = 25, T = 2$

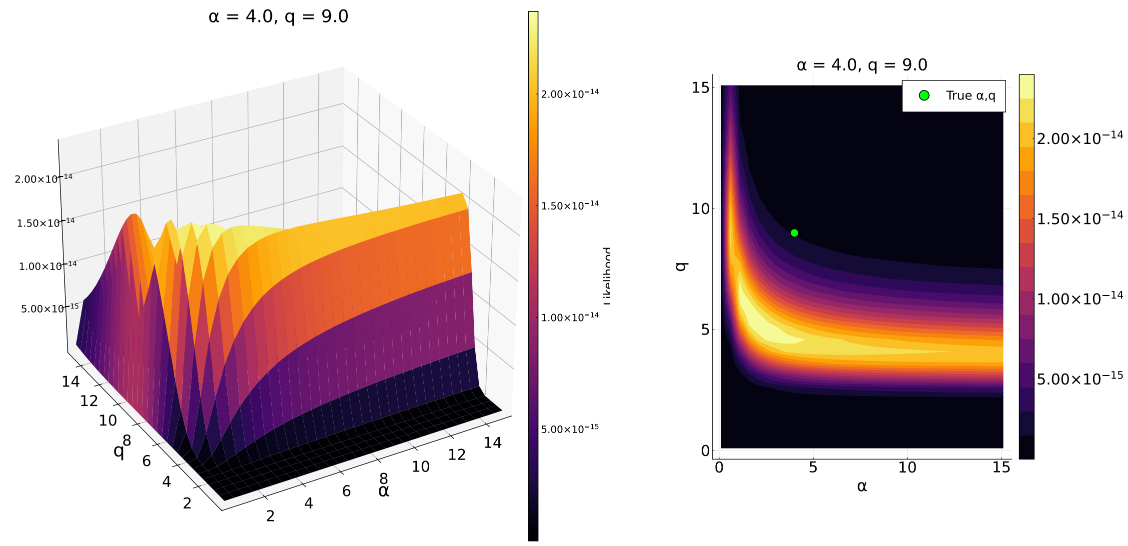

Figure 3: Likelihood function for data simulated by $\pi \sim Beta(4,1), N = 100, n = 25, T = 2$