Discrete Framework

November 5, 2018

Contents

1 One-population

1.1 Problem and assumptions

Probability distributions:

$$p(\theta) = 1$$

$$\begin{aligned} p(r_i|\theta) &= \frac{(g_i(\theta)\tau)^{r_i}}{r_i!} e^{-g_i(\theta)\tau} \\ p(r_1, \ldots r_M|\theta) &= \prod_{j=1}^{M} \frac{(g_j(\theta)\tau)^{r_j}}{r_j!} e^{-g_j(\theta)\tau} \end{aligned}$$

where:

- θ: stimulus on the circle $\mathcal{C}=$[0, 1);
- f : the tuning curve defined on the circle (can be considered as a periodic function);
- $g(\theta) = \int_0^1 f(\theta - y)s(y)dy$: the firing rate (can be considered as a periodic function);
- s: the convolution kernel, $\int_0^1 s(\theta)d\theta = 1$;
- M: the number of neurons;
- r_i: the number of spikes of the i-th neuron in a given time τ, conditionally independent Poissson r.v.;
- τ: the time for stimulus to be active; we always set $\tau = 1$.

Optimization Problem:

$$\begin{gathered}\max_{f_1(\cdot),\cdots,f_M(\cdot)} I(\mathbf{r};\theta)\\ f_- \leq f_i(\cdot) \leq f_+\\ \int g_i(\theta)p(\theta)d\theta = \bar{f}\end{gathered}$$

where f_-, f_+, $\bar{f}$ are positive constants. Note that when $p(\theta)=1$, $\int g_i(\theta)d\theta = \bar{f}$ is equivalent to $\int f_i(\theta)d\theta = \bar{f}$:

$$\bar{f} = \int g_i(\theta)d\theta = \int\int f_i(\gamma)s(\theta-\gamma)d\gamma d\theta = \int f_i(\gamma)\int s(\theta-\gamma)d\theta d\gamma = \int f_i(\gamma)d\gamma$$

Therefore the problem we consider:

$$\begin{gathered}\max_{f_1(\cdot),\cdots,f_M(\cdot)} I(\mathbf{r};\theta)\\ f_- \leq f_i(\cdot) \leq f_+\\ \int f_i(\theta)d\theta = \bar{f}\end{gathered} \tag{1}$$

Discretization and Cyclic Invariance:

Discretize $\theta \in [0,1)$ to be $\theta_i \triangleq \frac{i}{M}$, $f_0(\theta_i) \triangleq f_{-i}$, $g_0(\theta_i) \triangleq g_{-i}$, $i = 0, ..., M-1$. Here the number of bins M is equal to the number of neruons.

The convolution can be written as:

$$g_i = \sum_{j=0}^{M-1} f_{i-j}s_j = \sum_{j=0}^{M-1} f_j s_{i-j}$$

Assume the tuning curves are rotationally invariant, i.e. different f_i's are the translations (rotations in $[0,1)$) of the same function:

$$f_j(\theta) \triangleq f_0(\theta - \theta_j)$$

Therefore,

$$f_i(\theta_j) = f_0(\theta_j - \theta_i) = f_{i-j}$$

Similarily, g's are rotationally invariant:

$$g_i(\theta_j) = \int f_i(\gamma)s(\theta_j-\gamma)d\gamma = \int f_0(\gamma-\theta_i)s(\theta_j-\gamma)d\gamma = g_0(\theta_j-\theta_i) = g_{i-j}$$

Note:

- Because of periodicity of f, $(i-j)$ actually equals $(i-j) \mod M$.
- The cyclic invariance notation is consistent with Lorenzo's notes where $f_i(0) = f_i$.

Mutual Information:

$$\begin{aligned}I(\mathbf{r};\theta) &= D_{KL}(p(\mathbf{r},\theta)||p(\mathbf{r})p(\theta))\\ &= \int_\theta\int_{\mathbf{r}} p(\mathbf{r},\theta)\ln\left(\frac{p(\mathbf{r},\theta)}{p(\mathbf{r})p(\theta)}\right)d\mathbf{r}d\theta\\ &= \int_\theta\int_{\mathbf{r}} p(\mathbf{r}|\theta)p(\theta)\ln\left(\frac{p(\mathbf{r}|\theta)}{p(\mathbf{r})}\right)d\mathbf{r}d\theta\\ &= \sum_{i=1}^{M}\int_{\mathbf{r}} p(\mathbf{r}|\theta=\theta_i)\frac{1}{M}\ln\left(\frac{p(\mathbf{r}|\theta=\theta_i)}{p(\mathbf{r})}\right)d\mathbf{r}\\ &= \frac{1}{M}\sum_{i=1}^{M} D_{KL}\left(p(\mathbf{r}|\theta=\theta_i)\|p(\mathbf{r})\right)\\ &= D_{KL}\left(p(\mathbf{r}|\theta=0)\|p(\mathbf{r})\right)\end{aligned} \tag{2}$$

Note that the equation (2) follows from a conclusion

$$D_{KL}\left(p(\mathbf{r}|\theta=\theta_k)\|p(\mathbf{r})\right) = D_{KL}\left(p(\mathbf{r}|\theta=\theta_0)\|p(\mathbf{r})\right), \forall k = 0, 1, ..., M-1 \tag{3}$$

derived from rotational invariance. This conclusion is proved as a special case of equation (18) in Section 3. Applying this conclusion, we obtain the expression of Mutual Information as an expectation:

$$I(\mathbf{r};\theta) = E_{\mathbf{r}|\theta=0}\left[-\ln\left(\frac{p(\mathbf{r})}{p(\mathbf{r}|\theta=0)}\right)\right] = E_{\mathbf{r}|\theta=0}\left[-\ln\left(S(\mathbf{r})\right)\right] \tag{4}$$

$$S(\mathbf{r}) := \frac{p(\mathbf{r})}{p(\mathbf{r}|\theta=0)} = \frac{1}{M}\sum_{i=1}^{M}\prod_{k=1}^{M} {g_k}^{r_{i+k}-r_k} \tag{5}$$

Therefore we have the discretized version of the optimization problem:

$$\begin{array}{c} \max_{\{f_i\}_{i=1,\dots,M}} \left[-E_{\mathbf{r}|\theta=0}\ln\left(S(\mathbf{r})\right)\right] \\ f_- \leq f_i \leq f_+ \\ \frac{1}{M}\sum_{i=1}^{M} f_i = \bar{f} \end{array} \tag{6}$$

1.2 Gradient and Hessian

Reference: Lorenzo's notes. Notice that in there, the derivatives of $-I(\mathbf{r};\theta) = E_{\mathbf{r}|\theta=0}\ln\left(S(\mathbf{r})\right)$ are derived.

Summarized below:

- The partial derivative of $P := p(\mathbf{r}|\theta=0) = \prod_{j=1}^{M} \frac{{g_j}^{r_j}}{r_j!}e^{-g_j}$ with respect to g_i :

$$\begin{aligned} \frac{\partial P}{\partial g_i} &= \left(\frac{r_i}{g_i}-1\right)P \\ \frac{\partial^2 P}{\partial g_i \partial g_j} &= \left[\left(\frac{r_i}{g_i}-1\right)\left(\frac{r_j}{g_j}-1\right) + 1_{\{i=j\}}\left(-\frac{r_i}{g_i^2}\right)\right]P \end{aligned}$$

- The 1st derivatives of $I(\mathbf{r};\theta)$:

$$\begin{aligned} \frac{\partial I}{\partial g_i} &= -\sum_{\mathbf{r}}(\partial_i P)\ln(S) \\ &= -\sum_{\mathbf{r}} P\cdot\left(\frac{r_i}{g_i}-1\right)\ln(S) \\ &= E_{\mathbf{r}|\theta=0}\left[\left(1-\frac{r_i}{g_i}\right)\ln(S)\right] \end{aligned} \tag{7}$$

$$\frac{\partial I}{\partial f_i} = \sum_{k=1}^{M}\frac{\partial I}{\partial g_k}s_{k-i} \tag{8}$$

- The 2nd derivatives of $I(\mathbf{r};\theta)$:

$$\begin{aligned} \frac{\partial^2 I}{\partial g_i \partial g_j} &= -\sum_{\mathbf{r}}\left((\partial_{ij}^2 P)\ln(S) + (\partial_i P)\frac{\partial_j S}{S}\right) \\ &= -\sum_{\mathbf{r}}\left[\left(\left(\frac{r_i}{g_i}-1\right)\left(\frac{r_j}{g_j}-1\right) + 1_{\{i=j\}}\left(-\frac{r_i}{g_i^2}\right)\right)P\ln(S) + \frac{1}{g_i g_j}\sum_{\mathbf{r}}\frac{\sum_k r_i(r_{j+k}-r_j)\prod_l g_l^{r_{l+k}+r_l}}{\sum_k\prod_l g_l^{r_{k+l}}e^{g_l}r_l!}\right] \\ &= E_{\mathbf{r}|\theta=0}\left[-\left(\frac{r_i}{g_i}-1\right)\left(\frac{r_j}{g_j}-1\right)\ln(S) + \frac{1}{2g_i g_j}\frac{\sum_k(r_i-r_{i+k})(r_j-r_{j+k})\prod_l g_l^{r_{l+k}}}{\sum_k\prod_l g_l^{r_{k+l}}} + 1_{\{i=j\}}\cdot\left(\frac{r_i}{g_i^2}\right)\ln(S)\right] \quad (9) \\ &= E_{\mathbf{r}|\theta=0}\left[-\left(\frac{r_i}{g_i}-1\right)\left(\frac{r_j}{g_j}-1\right)\ln(S) + \frac{1}{g_i g_j}\frac{\sum_k r_i(r_j-r_{j+k})\prod_l g_l^{r_{l+k}}}{\sum_k\prod_l g_l^{r_{k+l}}} + 1_{\{i=j\}}\cdot\left(\frac{r_i}{g_i^2}\right)\ln(S)\right] \quad (10) \end{aligned}$$

$$\frac{\partial^2 I}{\partial f_i \partial f_j} = \sum_{k=1}^{M}\sum_{l=1}^{M}\frac{\partial^2 I}{\partial g_k \partial g_l}s_{k-i}s_{l-j} \tag{11}$$

2 Multi-populations

2.1 Problem and assumptions

Probability distributions:

Assume different populations are conditionally independent on θ:

$$P(\mathbf{r}|\theta) = \prod_{n=1}^{N}\prod_{j=1}^{M} P(r_{n,j}|\theta) = \prod_{n=1}^{N}\prod_{j=1}^{M} \frac{(g_{n,j}(\theta)\tau)^{r_{n,j}}}{r_{n,j}!} e^{-g_{n,j}(\theta)\tau}$$

where:

- M: the number of neurons in each population;
- N : the number of populations;
- Denote $\mathbf{r} = (r_{n,k})$, $n = 1, \ldots, N$, $k = 1, \ldots, M$;
- Same stimulus for all populations: $p(\theta) = 1$ uniform on the circle $[0, 1)$;
- $f_{n,j}$: the firing rate of the j-th neuron in the n-th population;
- Same convolution kernel for all populations: $g_n(\theta) = \int_0^1 f_n(\theta - y)s(y)dy$;
- Always set $\tau = 1$.

From now on, for convenience we **do not distinguish** f and its convolution g. It only affects the formulas of gradients which follow the same relations as in (8), (11).

Discretization and Cyclic Invariance:

Similar to the one-population case, $f_{n,0}(\theta_i) \triangleq f_{n,-i}$,

$$f_{n,i}(\theta_j) = f_{n,0}(\theta_j - \theta_i) = f_{n,i-j}$$

Mutual Information:

$$p(\mathbf{r}) = \sum_{i=1}^{M} p(\theta_i)p(\mathbf{r}|\theta = \theta_i) = \frac{1}{M}\sum_{i=1}^{M}\prod_{n=1}^{N}\prod_{k=1}^{M} \frac{{f_{n,k-i}}^{r_{n,k}}}{r_{n,k}!} e^{-f_{n,k-i}}$$

$$\frac{p(\mathbf{r})}{p(\mathbf{r}|\theta = 0)} = \frac{1}{M}\sum_{i=1}^{M}\prod_{n=1}^{N}\prod_{k=1}^{M} \left(\frac{f_{n,k-i}}{f_{n,k}}\right)^{r_{j,k}} = \frac{1}{M}\sum_{i=1}^{M}\prod_{n=1}^{N}\prod_{k=1}^{M} f_{j,k}^{r_{n,k+i}-r_{n,k}}$$

$$I(\mathbf{r};\theta) = E_{\mathbf{r}|\theta=0}\left[-\ln\left(\frac{p(\mathbf{r})}{p(\mathbf{r}|\theta = 0)}\right)\right] \tag{12}$$

Optimization Problem:

$$\begin{aligned} &\max_{\{f_{n,j}\}_{n=1,..,N}^{j=1,\ldots,M}} E_{\mathbf{r}|\theta=0}\left[-\ln\left(S(\mathbf{r})\right)\right] \\ &0 < (f_-)_n \le f_{n,j} \le (f_+)_n, \qquad n = 1, \ldots, N \\ &\tfrac{1}{M}\textstyle\sum_{j=1}^{M} f_{n,j} = \bar{f}_n, \qquad n = 1, \ldots, N \end{aligned} \tag{13}$$

where $(r_{j,k}|\theta = 0) \sim Poisson(f_{j,k})$,

$$S(\mathbf{r}) := \frac{1}{M}\sum_{i=1}^{M}\prod_{n=1}^{N}\prod_{k=1}^{M} f_{j,k}^{r_{n,k+i}-r_{n,k}} = \frac{1}{M}\sum_{i=1}^{M} Q^i(\mathbf{r}) \tag{14}$$

$$Q^i(\mathbf{r}) := \prod_{n=1}^{N}\prod_{k=1}^{M} f_{j,k}^{r_{n,k+i}-r_{n,k}} \tag{15}$$

2.2 Gradient and Hessian

Denote

$$P := p(\mathbf{r}|\theta = 0) = \prod_{n=1}^{N}\prod_{k=1}^{M} \frac{{f_{n,k}}^{r_{n,k}}}{r_{n,k}!} e^{-f_{n,k}}$$

the partial derivatives of P are:

$$\begin{aligned}
\frac{\partial P}{\partial f_{n,k}} &= \left(\frac{r_{n,k}}{f_{n,k}} - 1\right) P \\
\frac{\partial^2 P}{\partial f_{n,i}\partial f_{l,j}} &= \left[\left(\frac{r_{n,i}}{f_{n,i}} - 1\right)\left(\frac{r_{l,j}}{f_{l,j}} - 1\right) + 1_{\{(n,i)\neq(l,j)\}}\left(-\frac{r_{n,i}}{f_{n,i}^2}\right)\right] P
\end{aligned}$$

- The 1st derivatives of $I(\mathbf{r};\theta)$: similar to the arguments in Lorenzo's notes, the term $\sum_{\mathbf{r}} \frac{P}{S}\frac{\partial S}{\partial f_{n,k}}$ is zero, and thus

$$\begin{aligned}
\frac{\partial I}{\partial f_{n,k}} &= -\sum_{\mathbf{r}} \frac{P}{S}\frac{\partial S}{\partial f_{n,k}} - \sum_{\mathbf{r}} \frac{\partial P}{\partial f_{n,k}} \ln(S) \\
&= E_{\mathbf{r}|\theta=0}\left[-\left(\frac{r_{n,k}}{f_{n,k}} - 1\right)\ln(S)\right] \qquad (16)
\end{aligned}$$

where $n \in [N]$, $k \in [M]$.

- The 2nd derivatives of $I(\mathbf{r};\theta)$:

$$\begin{aligned}
\frac{\partial^2 I}{\partial f_{p,i}\partial f_{q,j}} &= E_{\mathbf{r}|\theta}\Bigg[-\left(\frac{r_{p,i}}{f_{p,i}} - 1\right)\left(\frac{r_{q,j}}{f_{q,j}} - 1\right)\ln(S) + \frac{\sum_{k=1}^{M}(r_{p,i+k} - r_{p,i})(r_{q,j+k} - r_{q,j})\prod_{l,m} f_{l,m}^{r_{l,m+k}}}{2 f_{p,i} f_{q,j} \sum_{k=1}^{M}\prod_{l,m} f_{l,m}^{r_{l,m+k}}} + 1_{(p,i)}^{(q,j)}\left(\frac{r_{p,i}}{f_{p,i}^2}\right)\ln(S)\Bigg] \\
&= E_{\mathbf{r}|\theta}\Bigg[-\left(\frac{r_{p,i}}{f_{p,i}} - 1\right)\left(\frac{r_{q,j}}{f_{q,j}} - 1\right)\ln(S) - \frac{r_{p,i}}{f_{p,i} f_{q,j}}\frac{\sum_{k=1}^{M}(r_{q,j+k} - r_{q,j})\prod_{l,m} f_{l,m}^{r_{l,m+k}}}{\sum_{k=1}^{M}\prod_{l,m} f_{l,m}^{r_{l,m+k}}} + 1_{(p,i)}^{(q,j)}\left(\frac{r_{p,i}}{f_{p,i}^2}\right)\ln(S)\Bigg] \qquad (17)
\end{aligned}$$

where $p, q \in [N]$, $i, j \in [M]$.

3 Neurons centered on a subset of positions

3.1 Problem and assumptions

Assume the neurons are centered on equi-distant positions on the circle:

- N: the number of centers=the number of neurons
- δ: the distance between centers (integer);
- M: the number of positions on the circle $[0, 1)$, $M = N\delta$;
- c_k: the position of the k-th center, $c_k = \delta k$, $k = 1, ..., N$;
- r_k: the number of spikes at the k-th center in a given time $\tau = 1$;
- f_{c_k} : the tuning curve of the k-th center on the circle;
- $g(\theta) = \int_0^1 f(\theta - y)s(y)dy$: the firing rate; we do not distinguish between f and g from now on.

Probability distributions:

$$p(\theta) = 1$$

$$\begin{aligned}
p(r_k|\theta) &= \frac{(f_{c_k}(\theta)\tau)^{r_k}}{r_k!} e^{-f_{c_k}(\theta)\tau} \\
p(r_1, \ldots r_N|\theta) &= \prod_{k=1}^{N} \frac{(f_{c_k}(\theta)\tau)^{r_k}}{r_k!} e^{-f_{c_k}(\theta)\tau}
\end{aligned}$$

Discretization and Cyclic Invariance:

$$\begin{aligned}\theta_i &:= \frac{i}{M}\\ f_i &:= f_0(\theta_i)\\ f_{c_k}(\theta) &= f_0(\theta - c_k)\\ f_k(\theta_m) &= f_0(\theta_m - c_k) = f_{m-c_k}\end{aligned}$$

Therefore $r_k|\theta_m$ satiesfies Poisson distribution:

$$p(r_1, \dots r_N|\theta_m) = \prod_{k=1}^{N} \frac{f_{m-c_k}^{r_k}}{r_k!} e^{-f_{m-c_k}}$$

Mutual Information:

First we show the following conclusion:

$$D_{KL}\left(p(\mathbf{r}|\theta_m)\|p(\mathbf{r})\right) = D_{KL}\left(p(\mathbf{r}|\theta_{m+\delta})\|p(\mathbf{r})\right) \tag{18}$$

Proof.

$$\frac{p(\mathbf{r})}{p(\mathbf{r}|\theta_m)} = \frac{\frac{1}{M}\sum_{i=1}^{M}\prod_{k=1}^{N}\frac{f_{i-c_k}^{r_k}}{r_k!}e^{-f_{i-c_k}}}{\prod_{k=1}^{N}\frac{f_{m-c_k}^{r_k}}{r_k!}e^{-f_{m-c_k}}} = \frac{1}{M}\sum_{i=1}^{M}\prod_{k=1}^{N}\left(\frac{f_{i-c_k}}{f_{m-c_k}}\right)^{r_k} e^{-(f_{i-c_k}-f_{m-c_k})}$$

$$\begin{aligned}D_{KL}\left(p(\mathbf{r}|\theta_m)\|p(\mathbf{r})\right) &= -E_{\mathbf{r}|\theta_m}\left[\ln\left(\frac{p(\mathbf{r})}{p(\mathbf{r}|\theta_m)}\right)\right]\\ &= -\sum_{\mathbf{r}}\prod_{k=1}^{N}\frac{f_{m-c_k}^{r_k}}{r_k!}e^{-f_{m-c_k}}\ln\left(\frac{1}{M}\sum_{i=1}^{M}\prod_{k=1}^{N}\left(\frac{f_{i-c_k}}{f_{m-c_k}}\right)^{r_k}e^{-(f_{i-c_k}-f_{m-c_k})}\right)\\ &= -\sum_{\mathbf{r}}C_1(\mathbf{r})\prod_{k=1}^{N}f_{m-c_k}^{r_k}e^{-f_{m-c_k}}\ln\left(\prod_{k=1}^{N}f_{m-c_k}^{-r_k}e^{f_{m-c_k}}\cdot C_2(\mathbf{r})\right)\end{aligned}$$

where $C_1(\mathbf{r}) = \frac{1}{\prod_{k=1}^{M} r_k!}$, $C_2(\mathbf{r}) = \frac{1}{M}\sum_{i=1}^{M}\prod_{k=1}^{N} f_{i-c_k}^{r_k}e^{-f_{i-c_k}}$ does not depend on m. Since $c_k = k\delta$, $m+\delta-c_k = m - c_{k-1}$,

$$\begin{aligned}D_{KL}\left(p(\mathbf{r}|\theta_{m+\delta})\|p(\mathbf{r})\right) &= -\sum_{\mathbf{r}}C_1(\mathbf{r})\prod_{k=1}^{N}f_{m-c_{k-1}}^{r_k}e^{-f_{m-c_{k-1}}}\ln\left(\prod_{k=1}^{N}f_{m-c_{k-1}}^{-r_k}e^{f_{m-c_{k-1}}}\cdot C_2(\mathbf{r})\right)\\ &= -\sum_{\mathbf{r}}C_1(\mathbf{r})\prod_{k=1}^{N}f_{m-c_k}^{r_{k+1}}e^{-f_{m-c_k}}\ln\left(\prod_{k=1}^{N}f_{m-c_k}^{-r_{k+1}}e^{f_{m-c_k}}\cdot C_2(\mathbf{r})\right)\end{aligned}$$

Taking the cyclic permutation of $\mathbf{r}$ by 1 such that $\tilde{r}_k = r_{k+1}$. One can easily verify that $C_1(\mathbf{r})$ and $C_2(\mathbf{r})$ are invariant under cyclic permutations of $\mathbf{r}$. Therefore, we obtain

$$\begin{aligned}D_{KL}\left(p(\mathbf{r}|\theta_{m+\delta})\|p(\mathbf{r})\right) &= \sum_{\mathbf{r}}C_1(\mathbf{r})\prod_{k=1}^{N}f_{m-c_k}^{r_{k+1}}e^{-f_{m-c_k}}\ln\left(\prod_{k=1}^{N}f_{m-c_k}^{-r_{k+1}}e^{f_{m-c_k}}\cdot C_2(\mathbf{r})\right)\\ &= \sum_{\tilde{\mathbf{r}}}C_1(\tilde{\mathbf{r}}_{-1})\prod_{k=1}^{N}f_{m-c_k}^{\tilde{r}_k}e^{-f_{m-c_k}}\ln\left(\prod_{k=1}^{N}f_{m-c_k}^{-\tilde{r}_k}e^{f_{m-c_k}}\cdot C_2(\tilde{\mathbf{r}}_{-1})\right)\\ &= \sum_{\tilde{\mathbf{r}}}C_1(\tilde{\mathbf{r}})\prod_{k=1}^{N}f_{m-c_k}^{\tilde{r}_k}e^{-f_{m-c_k}}\ln\left(\prod_{k=1}^{N}f_{m-c_k}^{-\tilde{r}_k}e^{f_{m-c_k}}\cdot C_2(\tilde{\mathbf{r}})\right)\\ &= D_{KL}\left(p(\mathbf{r}|\theta_m)\|p(\mathbf{r})\right)\end{aligned}$$

Note that a special case is when $\delta = 1$, then $D_{KL}\left(p(\mathbf{r}|\theta_{m+1})\|p(\mathbf{r})\right) = D_{KL}\left(p(\mathbf{r}|\theta_m)\|p(\mathbf{r})\right) = D_{KL}\left(p(\mathbf{r}|\theta_0)\|p(\mathbf{r})\right)$ for all m, which gives our conclusion (2) in Section 1. $\square$

Now we derive the mutual information:

$$
\begin{aligned}
I(\mathbf{r};\theta) &= D_{KL}(p(\mathbf{r},\theta)||p(\mathbf{r})p(\theta)) \\
&= \int_\theta \int_\mathbf{r} p(\mathbf{r},\theta) \ln\left(\frac{p(\mathbf{r},\theta)}{p(\mathbf{r})p(\theta)}\right) d\mathbf{r} d\theta \\
&= \int_\theta \int_\mathbf{r} p(\mathbf{r}|\theta)p(\theta) \ln\left(\frac{p(\mathbf{r}|\theta)}{p(\mathbf{r})}\right) d\mathbf{r} d\theta \\
&= \frac{1}{M}\sum_{m=1}^{M} D_{KL}\left(p(\mathbf{r}|\theta_m)\|p(\mathbf{r})\right) \\
&= \frac{1}{\delta}\sum_{m=1}^{\delta} D_{KL}\left(p(\mathbf{r}|\theta_m)\|p(\mathbf{r})\right) \\
&= \frac{1}{M}\sum_{m=1}^{M} E_{\mathbf{r}|\theta_m}\left[-\ln\left(\frac{p(\mathbf{r})}{p(\mathbf{r}|\theta_m)}\right)\right] \qquad (19) \\
&= \frac{1}{\delta}\sum_{m=1}^{\delta} E_{\mathbf{r}|\theta_m}\left[-\ln\left(\frac{p(\mathbf{r})}{p(\mathbf{r}|\theta_m)}\right)\right] \qquad (20)
\end{aligned}
$$

Optimization Problem:

$$
\begin{gathered}
\max_{\{f_j\}_{j=1,..,M}} \tfrac{1}{\delta}\textstyle\sum_{m=1}^{\delta} E_{\mathbf{r}|\theta_m}\left[-\ln\left(S_m(\mathbf{r})\right)\right] \\
0 < f_- \leq f_j \leq f_+ \\
\tfrac{1}{M}\textstyle\sum_{j=1}^{M} f_j = \bar{f}
\end{gathered} \qquad (21)
$$

where

$$
S_m(\mathbf{r}) := \frac{p(\mathbf{r})}{p(\mathbf{r}|\theta_m)} = \frac{1}{M}\sum_{i=1}^{M}\prod_{k=1}^{N}\left(\frac{f_{i-c_k}}{f_{m-c_k}}\right)^{r_k} e^{-(f_{i-c_k}-f_{m-c_k})}
$$

3.2 Gradient

For simplicity, we introduce the following notations:

$$
\begin{aligned}
P_m(\mathbf{r}) &:= P(\mathbf{r}|\theta_m) = \prod_{k=1}^{N}\frac{f_{m-c_k}^{r_k}}{r_k!}e^{-f_{m-c_k}} \\
i \sim j &: (i-j) \mod \delta = 0
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
I(\mathbf{r};\theta) &= -\frac{1}{M}\sum_{m=1}^{M} E_{\mathbf{r}|\theta_m}\left[\ln\left(\frac{p(\mathbf{r})}{p(\mathbf{r}|\theta_m)}\right)\right] \\
&= -\frac{1}{M}\sum_{m=1}^{M}\sum_{\mathbf{r}} P_m(\mathbf{r})\left[\ln S_m(\mathbf{r})\right] \\
\frac{\partial I(\mathbf{r};\theta)}{\partial f_i} &= -\frac{1}{M}\sum_{m=1}^{M}\sum_{\mathbf{r}}\left[\frac{\partial P_m(\mathbf{r})}{\partial f_i}\ln S_m(\mathbf{r}) + \frac{P_m(\mathbf{r})}{S_m(\mathbf{r})}\frac{\partial S_m(\mathbf{r})}{\partial f_i}\right] \qquad (22)
\end{aligned}
$$

First, we compute the partial derivatives of $P_m(\mathbf{r}) = \prod_{k=1}^{N} \frac{f_{m-c_k}^{r_k}}{r_k!} e^{-f_{m-c_k}}$. Notice that it only contains certain f_i's such that $i \sim m$:

$$\frac{\partial P_m}{\partial f_i} = 1_{\{i\sim m\}} \left(\frac{r_{\hat{m}}}{f_i} - 1 \right) P_m$$

where $\hat{m} := \frac{m-i}{\delta}$. Next, we show that the second term in equation (22) is zero (this conclusion is also used in deriving the gradient in Section 1):

$$\begin{aligned}
\frac{\partial S_m}{\partial f_i} &= \frac{1}{M} \sum_{j=1}^{M} \frac{\partial}{\partial f_i} \left[\frac{\prod_{k=1}^{N} f_{j-\delta k}^{r_k} e^{-f_{j-\delta k}}}{\prod_{k=1}^{N} f_{m-\delta k}^{r_k} e^{-f_{m-\delta k}}} \right] \\
&= \frac{1}{M} \sum_{j=1}^{M} \left(1_{\{i\sim j\}} \left(\frac{r_{\hat{j}}}{f_i} - 1 \right) + 1_{\{i\sim m\}} \left(-\frac{r_{\hat{m}}}{f_i} + 1 \right) \right) \frac{\prod_{k=1}^{N} f_{j-\delta k}^{r_k} e^{-f_{j-\delta k}}}{\prod_{k=1}^{N} f_{m-\delta k}^{r_k} e^{-f_{m-\delta k}}} \\
f_i \sum_{\mathbf{r}} \sum_{m=1}^{M} \frac{P_m(\mathbf{r})}{S_m(\mathbf{r})} \frac{\partial S_m(\mathbf{r})}{\partial f_i} &= \frac{1}{M} \sum_{\mathbf{r}} \sum_{m=1}^{M} \frac{P_m(\mathbf{r})}{S_m(\mathbf{r})} \sum_{j=1}^{M} \left[1_{\{i\sim j\}} \left(r_{\hat{j}} - f_i \right) - 1_{\{i\sim m\}} \left(r_{\hat{m}} - f_i \right) \right] \frac{P_j(\mathbf{r})}{P_m(\mathbf{r})} \\
&= \frac{1}{M} \sum_{\mathbf{r}} \sum_{m=1}^{M} \frac{1}{S_m(\mathbf{r})} \sum_{j=1}^{M} P_j(\mathbf{r}) \left[1_{\{i\sim j\}} \left(r_{\hat{j}} - f_i \right) - 1_{\{i\sim m\}} \left(r_{\hat{m}} - f_i \right) \right] \\
&= \sum_{\mathbf{r}} \sum_{m=1}^{M} \frac{\sum_{j=1}^{M} \prod_{k=1}^{N} \frac{f_{j-\delta k}^{r_k}}{r_k!} e^{-f_{j-\delta k}} \left[1_{\{i\sim j\}} \left(r_{\hat{j}} - f_i \right) - 1_{\{i\sim m\}} \left(r_{\hat{m}} - f_i \right) \right]}{\sum_{l=1}^{M} \prod_{k=1}^{N} \left(\frac{f_{l-\delta k}}{f_{m-\delta k}} \right)^{r_k} e^{-(f_{l-\delta k} - f_{m-\delta k})}} \\
&= \sum_{\mathbf{r}} \frac{\sum_{m=1}^{M} \sum_{j=1}^{M} \prod_{k=1}^{N} \left(f_{m-\delta k} f_{j-\delta k} \right)^{r_k} e^{-(f_{j-\delta k} + f_{m-\delta k})} \cdot 1_{\{i\sim j\}} \left(r_{\hat{j}} - f_i \right)}{\sum_{l=1}^{M} \prod_{k=1}^{N} f_{l-\delta k}^{r_k} r_k! e^{-f_{l-\delta k}}} \\
&\quad - \sum_{\mathbf{r}} \frac{\sum_{m=1}^{M} \sum_{j=1}^{M} \prod_{k=1}^{N} \left(f_{m-\delta k} f_{j-\delta k} \right)^{r_k} e^{-(f_{j-\delta k} + f_{m-\delta k})} \cdot 1_{\{i\sim m\}} \left(r_{\hat{m}} - f_i \right)}{\sum_{l=1}^{M} \prod_{k=1}^{N} f_{l-\delta k}^{r_k} r_k! e^{-f_{l-\delta k}}} \\
&= 0
\end{aligned}$$

Thus

$$\begin{aligned}
\frac{\partial I(\mathbf{r};\theta)}{\partial f_i} &= -\frac{1}{M} \sum_{m=1}^{M} \sum_{\mathbf{r}} \frac{\partial P_m(\mathbf{r})}{\partial f_i} \ln S_m(\mathbf{r}) \\
&= -\frac{1}{\delta} \sum_{m=1}^{\delta} \sum_{\mathbf{r}} \frac{\partial P_m(\mathbf{r})}{\partial f_i} \ln S_m(\mathbf{r}) \\
&= -\frac{1}{\delta} \sum_{m=1}^{\delta} \sum_{\mathbf{r}} 1_{\{i\sim m\}} \left(\frac{r_{\hat{m}}}{f_i} - 1 \right) P_m(\mathbf{r}) \ln S_m(\mathbf{r}) \\
&= \frac{1}{\delta} \sum_{m=1}^{\delta} E_{\mathbf{r}|\theta_m} \left[1_{\{i\sim m\}} \left(1 - \frac{r_{\hat{m}}}{f_i} \right) \ln S_m(\mathbf{r}) \right] \\
&= \frac{1}{\delta} \sum_{m=1}^{\delta} E_{\mathbf{r}|\theta_m} \left[\sum_{k=1}^{N} 1_{\{i=m+\delta k\}} \left(1 - \frac{r_k}{f_{m+\delta k}} \right) \ln S_m(\mathbf{r}) \right]
\end{aligned} \tag{23}$$

where the second equality can be shown in a similar way as in the proof of (18).