Statistical mechanics of spike analysis model by use of log-linear model

Hayaru Shouno† ¹ and Masato Okada‡
† Faculty of Engineering, Yamaguchi University, Ube 755-8611, Japan
‡ Graduate School of New Frontier, University of Tokyo, Kashiwa, 277-8561, Japan
‡ Japan Scientific Technology Corp.
‡ RIKEN BSI, Wako, 351-0198, Japan

In this paper, we consider the mathematical framework for extracting higher-order correlation of neural firings using a log-linear model. From statistical-mechanical point of view, the log-linear model can be regarded as a multi-body interacted Ising spin model or the Boltzman machine with higher-order interactions. The estimation of higher-order correlation of neural firing corresponds to that of higher-order interactions in this Ising spin system, and to the hyper-parameter estimation in the Bayesian inference.

We assume \( N \) neurons are observed and each neuron can have two states, that is, one is firing state denoted as ‘+1’ and the other is non-firing state denoted as ‘−1’. We describe the state of neurons as \( \xi \), and each component, \( \xi_i \in \{-1, +1\} \) where \( 1 \leq i \leq N \), describe the single neuron state. Then the log-linear model can be described as

\[
P(\xi; J) = \frac{1}{Z_\pi(J)} \exp \left( \sum_{r=1}^{R} \sum_{i_1 < \cdots < i_r} J_{i_1 \cdots i_r} \xi_{i_1} \cdots \xi_{i_r} \right),
\]

where \( Z_\pi(J) = \text{Tr}_\xi \exp(\sum_{r=1}^{R} \sum_{i_1 < \cdots < i_r} J_{i_1 \cdots i_r} \xi_{i_1} \cdots \xi_{i_r}) \), \( J_{i_1 \cdots i_r} \) are hyper-parameters which can be regarded as the multi-body interaction strengths. Since the number of hyper-parameters is \( 2^N - 1 \), carrying out whole hyper-parameters estimation is not practical. Thus, we assumed those hyper-parameters obey the multi-dimensional normal distribution

\[
P(J_{i_1 \cdots i_r}) = \sqrt{\frac{N!}{\Delta_{r}^2}} \exp \left( -\frac{\Delta_{r}^2}{2} (J_{i_1 \cdots i_r} - J_{i_1 \cdots i_r}^M)^2 \right),
\]

and tried to estimate the means \( \{ \frac{J_{i_1 \cdots i_r}^M}{\Delta_{r}^2} \} \) and the variances \( \{ \frac{\Delta_{r}^2}{N!} \} \).

Furthermore, in our model, we introduce observation noises like binary symmetric channel (BSC), that is, \( P(\tau = \pm 1|\xi = \pm 1) = p \) where \( \tau \) means the observation for the signal \( \xi \), and \( p \) means flip probability. Thus, the conditional probability \( P(\tau|\xi; \beta_p) \) can be described as

\[
P(\tau|\xi; \beta_p) = \exp(\beta_p \sum_{i=1}^{N} \xi_i \tau_i) / \left( 2 \cosh(\beta_p) \right)^N
\]

where \( \beta_p = \frac{1}{2} \ln \frac{1-p}{p} \).

To estimate hyper-parameter \( \{ J_r \} \) and \( \{ \Delta_r^2 \} \), we applied maximization of log-likelihood which can be described as those partition functions

\[
\ln P(\tau; \beta_p, J) = \ln Z - \ln Z_\pi - \ln Z_L,
\]

where

\[
Z(J) = \text{Tr} \exp \left( \sum_{r=1}^{R} \sum_{i_1 < \cdots < i_r} J_{i_1 \cdots i_r} \sigma_{i_1} \cdots \sigma_{i_r} + \beta_p \sum_{i=1}^{N} \tau_i \sigma_i \right),
\]

and \( Z_L = (2\cosh(\beta_p))^N \).

These partition functions includes randomness of \( \{ J_{i_1 \cdots i_r} \} \), so that we applied replica method to evaluate avaraged those log-likelihoods. For \( \{ Z^0_n \} \), we obtain

\[
f_n = \sum_r J_r (r-1)m_n^r - \sum_r \frac{\Delta_n^2}{4}(1 - q_n^r - q_n^{r-1} + q_n^{r+1}) - \int Du \ln 2\cosh G_n(u),
\]

where \( m_n = \int D\sigma \tanh G_n(u), q_n = \int D\sigma \tanh^2 G_n(u) \), and

\[
G_n(u) = G(u; m_n, q_n) = \sqrt{\sum_r \frac{\Delta_n^2}{2} q_n^{r-1} u} + \sum_r J_r m_n^{r-1}
\]

¹E-mail: shouno@yamaguchi-u.ac.jp
for log-likelihood. Thus we obtain the following algorithm:

$$f = \sum_{r} J_{r}(r-1)m^{r} - \sum_{r} \frac{\Delta^{2}}{4}(1 - q^{r} - rq^{-1} + r^{q}) - \frac{1}{N} \sum_{i=1}^{N} \int Du_{i} \ln 2\cosh(G(u_{i}) + \beta_{r}\tau_{i}),$$

(6)

where $m = \frac{1}{N} \sum_{i=1}^{N} \int Du_{i} \tanh(G(u_{i}) + \beta_{r}\tau_{i})$, $q = \frac{1}{N} \sum_{i=1}^{N} \int Du_{i} \tanh^{2}(G(u_{i}) + \beta_{r}\tau_{i})$, and $G(u_{i})$ is the same form to eq.(5) using $m$ and $q$, that is $G(u_{i}; m, q)$. Then we obtain the log-likelihood as $K = [\ln P(\tau)]_{f} = -N(f - f_{\tau} - \ln 2\cosh/\beta_{r}).$

In this study, to estimate hyper-parameters $\{J_{r}\}$ and $\{\Delta_{r}^{2}\}$, we adopted the gradient ascent method for log-likelihood. Thus we obtain the following algorithm:

$$J_{r}^{t+1} = J_{r}^{t} + \eta \frac{\partial J_{r}}{\partial \eta}, \quad \Delta_{r}^{2t+1} = \Delta_{r}^{2t} + \eta \frac{\partial \Delta_{r}^{2}}{\partial \Delta_{r}^{2}}.$$  

(7)

One purpose of our study is to evaluate the performance of the algorithm described as eq.(7). However, the algorithm depends on the observation $\{\tau_{i}\}$ which includes randomness. Hence, we should evaluate the averaged conduct of the algorithm over the observation $\{\tau_{i}\}$. To obtain the averaged log-likelihood $[K]_{\tau}$, we assumed the prior probability of can be described as $P(\xi) = \prod_{i=1}^{N} P(\xi_{i}) = \prod_{i=1}^{N} \int Dz_{i} \frac{\exp(\xi G_{i}(z_{i}))}{2\cosh(\xi G_{i})}$ in the limit of $N \to \infty$, where the superscript $t$ means that the true hyper-parameters $\{J_{r}\}$, $\{\Delta_{r}^{2}\}$ are given.

The observation process can be also described as the product of the distribution $P(\tau|\xi) = \prod_{i=1}^{N} P(\tau_{i}|\xi_{i}) = \prod_{i=1}^{N} \frac{\exp(\xi J_{r}(z_{i}))}{2\cosh(\xi J_{r})}$. Then the averaged log-likelihood can be described as

$[K]_{\tau} = [\ln Z]_{\tau} - \ln Z_{\tau} = -N(f - f_{\tau})$,  

(8)

$[f]_{\tau} = \sum_{r} J_{r}(r-1)m^{r} - \sum_{r} \frac{\Delta^{2}}{4}(1 - q^{r} - rq^{-1} + r^{q}) - \sum_{r} w(\tau) \int Du_{i} \ln 2\cosh(G(u_{i}) + \beta_{r}\tau_{i})$,  

(9)

where $w(\tau) = \int Dz \frac{\cosh(G_{i}(z_{i}) + \beta_{r}\tau)}{2\cosh(G_{i}(z_{i}) \cosh/\beta_{r})}$. The order parameters become $m = \sum_{\tau_{r}=\pm 1} w(\tau) \int Du_{i} \tanh(G(u_{i}) + \beta_{r}\tau)$ and $q = \sum_{\tau_{r}=\pm 1} w(\tau) \int Du_{i} \tanh^{2}(G(u_{i}) + \beta_{r}\tau)$.

Fig. 1 shows the trajectories of hyper-parameters estimation of $\{J_{r}\}$ for $\{J_{r}^{*}\} = (0.8, 0.2)$, $\{\Delta_{r}^{2}\} = (0.1, 0.1)$ where each trajectory starts with several initial states. The left figure shows the averaged trajectories of gradient-ascent algorithm, and the right one shows the simulation results corresponds to the left. The averaged hyper-parameters estimation does not converge to the true hyper-parameters, so that, we conclude it is difficult to determine the true hyper parameters from the log-linear model.

References