Information gain on variable neuronal firing rate

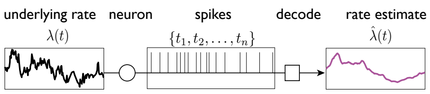
小山　慎介　　モデリング研究系　助教

Summary

The question of how much information can be theoretically gained from variable neuronal firing rate is investigated. We employ the statistical concept of information based on the Kullback-Leibler divergence, which is the key quantity for understanding the neural rate code.

- ✓ For rate-modulated renewal processes, the minimum information gain is derived, which is attained if and only if the ISI distribution is given by the gamma distribution.
- ✓ We give an explicit interpretation of the KL divergence in terms of detectability of rate variation: the lower bound of detectable rate variation, below which the temporal variation of firing rate is undetectable with a Bayesian decoder, is entirely determined by the KL divergence.
- ✓ The information gained from spike trains recored from a cortical area is estimated to be about 0.1 bits/spikes, with which roughly 4 and more trials are necessary to detect the underlying rate variations.

Neural Coding Problem



We assume that the firing rate carries a significant portion of the neural information.

- How much information about $\lambda(t)$ do spike trains carry?
- How well can the underlying rate $\lambda(t)$ be estimated from observed spike trains?

1. Spike Train Model

Rescaled renewal process:

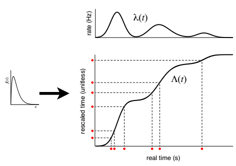
$$x_1, x_2, \ldots, x_n \overset{\text{i.i.d}}{\sim} f(x)$$

where

$$\int_0^\infty x f(x)dx = 1$$

By applying the time-rescaling transformation:

$$x_i = \Lambda(t_i) - \Lambda(t_{i-1}) := \int_{t_{i-1}}^{t_i} \lambda(t)dt$$



The probability density of $\{t_i\} := \{t_1, \ldots, t_n\}$ in the interval [0,*T*] is obtained as

$$p(\{t_i\}|\{\lambda(t)\}) = p(t_1|\{\lambda(t)\}) \cdot \prod_{i=2}^{n} \lambda(t_i) f(\Lambda(t_i) - \Lambda(t_{i-1})) \cdot P((t_n, T]|\{\lambda(t)\})$$

where $p(t_1|\{\lambda(t)\})$ is the density of the first spike occurring at t_1, and $P((t_n, T]|\{\lambda(t)\})$ is the probability of no spikes being observed in $(t_n, T]$.

2. Kullback-Leibler Divergence

In order to quantify the information about the rate variation, $\lambda(t)$ relative to the mean firing rate:

$$\mu = \frac{1}{T}\int_0^T \lambda(t)dt$$

we introduce the Kullback-Leibler (KL) divergence:

$$D[p(\cdot|\{\lambda(t)\})||p(\cdot|\mu)] = \frac{1}{\mu T}\sum_{n=0}^{\infty}\int_0^T\int_{t_1}^T \cdots \int_{t_{n-1}}^T p(\{t_i\}|\{\lambda(t)\}) \times \log\frac{p(\{t_i\}|\{\lambda(t)\})}{p(\{t_i\}|\mu)} dt_1 dt_2 \cdots dt_n \quad \text{[nats/spike]}$$

Note that the KL divergence is a functional of $\lambda(t)$ and $f(x)$.

Minimum KL Divergence with respect to *f*(*x*)

For a given rate process $\lambda(t)$, the minimum the KL divergence is given by

$$\min_f D[p(\cdot;\{\lambda(t)\})||p(\cdot;\mu)] = \frac{1}{C_V^2}\frac{1}{T}\int_0^T \frac{\lambda(t)}{\mu}\ln\frac{\lambda(t)}{\mu}dt + O\left(\frac{\sigma}{\tau\mu^2}\right). \quad \text{... (1)}$$

where σ, τ are the amplitude and characteristic timescale of $\lambda(t)$, respectively, and C_V is the coefficient of variation of *f*(*x*). Moreover, this minimum KL divergence is attained if and only if *f*(*x*) is the gamma distribution.

- ✓ The KL divergence increases as σ/μ is increased.
- ✓ The KL divergence increases as CV is decreased.
- ✓ The KL divergence significantly depends on the dispersion property of firing, which is described by *f*(*x*). Particularly, it attains the minimum for *f*(*x*) being the gamma distribution.

3. Bayesian Decoding

The posterior probability of the firing rate $\lambda(t)$ given the spike trains $\{t_i\}$ is obtained by the Bayes' theorem as

$$p(\{\lambda(t)\}|\{t_i\},\gamma) = \frac{p(\{t_i\}|\{\lambda(t)\})p(\{\lambda(t)\}|\gamma)}{p(\{t_i\}|\gamma)}$$

We choose the prior distribution of the firing rate such that the large gradient of $\lambda(t)$ is penalized with

$$p(\{\lambda(t)\}|\gamma) = \frac{1}{Z(\gamma)}\exp\left[-\frac{1}{2\gamma^2}\int_0^T\left(\frac{d\lambda(t)}{dt}\right)^2 dt\right]$$

where the hyperparameter γ controls the roughness of the time-dependent rate $\lambda(t)$; with small value of γ, the model requires the constant firing rate and vice versa.
The optimal hyperparameter γ is determined by maximizing the marginal likelihood:

$$p(\{t_i\}|\gamma) = \int p(\{t_i\}|\{\lambda(t)\})p(\{\lambda(t)\}|\gamma)\mathcal{D}\{\lambda(t)\}$$

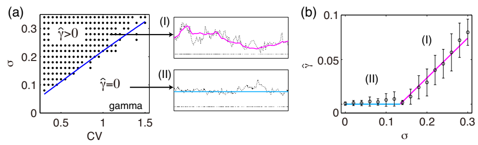
where the integration over the path space can be computed by the path integral method.

Detectability of Rate variation

The underlying rate variation is detectable if

$$D[p(\cdot|\{\lambda(t)\})||p(\cdot|\mu)] > \frac{\phi(0)}{4\mu\int_0^\infty \phi(u)du} \quad \text{... (2)}$$

where $\phi(u) = \langle(\lambda(t)-\mu)(\lambda(t+u)-\mu)\rangle$ is the correlation function of $\lambda(t)$.



For instance, if a spike is, on average, expected to be observed in the characteristic timescale of the rate variation, it is necessary for the spike train to carry more than 0.25 nats (=0.36 bits) per spike information so that the underlying rate variation is detectable.

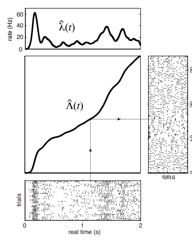
4. Data Analysis

We analyzed the biological spike data recorded from 216 neurons in the visual cortical area MT of three adult rhesus macaques, which is publicly available from Neural Signal Archive (http://www.neuralsignal.org). The experimental details are found in Britten et al. (1992).

Method:

Let $\{t_1^j, \ldots, t_{n_j}^j\}_{j=1}^N$ be N identical and independent trials in the interval [0,*T*].

1) Estimate the firing rate $\hat{\lambda}(t)$ from $\{t_1^j, \ldots, t_{n_j}^j\}_{j=1}^N$ using a kernel density estimator (Shimazaki, 2010).
2) Apply the time-rescaling transformation $x_i^j = \int_{t_{i-1}^j}^{t_i^j}\hat{\lambda}(t)dt$ to obtain the rescaled ISIs $\{x_2^j, \ldots, x_{n_j}^j\}_{j=1}^N$.
3) Compute the CV from the rescaled ISIs.
4) Compute the lower bound of KL divergence by Eq.(1).



Result:

We analyzed 1,064 datasets that contained trials ≧10 and ISIs ≧1,000.

- The estimated (lower bound of) KL divergence was ≈0.1 bits/spike on average.
- The minimum information given by the r.h.s of Eq. (2), which is required for detecting the underlying rate variation, was estimated ≈0.4 bits/spike.

Thus, roughly 4 and more trials are necessary to detect the underlying rate variation.

References

[1] S. Koyama and L. Kostal. (2013) The effect of interspike interval statistics on the information gain under the rate coding hypothesis. *Mathematical Biosciences and Engineering*, to appear.

[2] S. Koyama. (2013) Coding efficiency and detectability of rate fluctuations with non-Poisson neuronal firing. *Advances in Neural Information Processing Systems*, **25**, to appear.

[3] S. Koyama, T. Omi, R. E. Kass and S. Shinomoto. (2013) Information transmission using non-Poisson regular firing. *Neural Computation*, **25**: 854-876.

[4] S. Koyama. (2012) Estimating neural firing rates: an empirical Bayes approach. *Lecture Notes in Computer Science*, **7664**: 51-59.