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 recorded 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.

### Detection of Rate variation

The underlying rate variation is detectable if and only if the ISI distribution is given by the gamma distribution. How well can the underlying rate \( \lambda(t) \) be estimated from observed spike trains?

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

### 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

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

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

\[
p(\{t_i\}|\gamma) = \int p(\{t_i\}|\lambda(t))p(\lambda(t)|\gamma)P(\lambda(t))
\]

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

### Minimum KL Divergence with respect to \( f(x) \)

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

\[
\min \mathcal{D}(p(\lambda(t))|\mu) = \frac{1}{2} \int \tau(t)^2 - 2\int \mu(t) dt + O(\frac{\sigma}{\tau})
\]

where \( \sigma \) and \( \tau \) are the amplitude and characteristic timescale of \( \lambda(t) \), respectively, and \( C_{\lambda} \) is the coefficient of variation of \( \lambda(t) \). Moreover, this minimum KL divergence is attained if and only if \( f(x) \) is the gamma distribution.

The KL divergence increases as \( \sigma \) or \( \tau \) increases.

The KL divergence increases as \( CV \) decreases.

The KL divergence significantly depends on the dispersion property of firing, which is described by \( CV \). Particularly, it attains the minimum for \( CV = 1 \) being the gamma distribution.

### References


---

**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 recorded 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**

Underlying rate neuron spikes decode rate estimate

![Neural Coding Problem Diagram](image)

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?

**Spike Train Model**

Rescaled renewal process:

\[ x_i, x_{i+1}, \ldots, x_n \sim f(x) \]

where

\[ \int f(x) dx = 1 \]

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

\[
p(t_i|\lambda(t)) = \prod_{i=1}^{n} p(t_i|\lambda(t)) = \prod_{i=1}^{n} p(t_i|\lambda(t_i-1)) = \prod_{i=1}^{n} p(t_i|\lambda(t))
\]

where \( p(t_i|\lambda(t)) \) is the density of the first spike occurring at \( t_i \) and \( p(t_i|\lambda(t)) \) is the probability of no spikes being observed in \([t_{i-1}, t_i]\).

**Kullback-Leibler Divergence**

In order to quantify the information about the rate variation, \( \lambda(t) \), relative to the mean firing rate, we introduce the Kullback-Leibler (KL) divergence:

\[
\mathcal{D}(p(\lambda(t))|\mu) = \frac{1}{2} \sum_{\lambda} \int \left( \lambda(t)^2 - 2\mu(t) \lambda(t) + \mu(t) \right) \frac{p(\lambda(t))}{p(\mu(t))} dt
\]

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

**Data Analysis**

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

**Method:**

Given \( \{t_1, \ldots, t_I\} \) be \( N \) independent and identical trials in the interval \([0,T]\):

1. Estimate the firing rate \( \hat{\lambda}(t) \) from \( \{t_1, \ldots, t_I\} \) using a kernel density estimator (Shimazaki, 2010).
2. Apply the time-rescaling transformation \( x_i = \int t_i \hat{\lambda}(t) dt \) to obtain the rescaled ISIs \( \{x_1, \ldots, x_I\} \).
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 \( n \leq 10 \) and ISIs \( n \leq 1,000 \).

- The estimated (lower bound) of KL divergence was \( \approx 0.1 \) 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 \( \approx 0.4 \) bits/spike.

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