8. Entropy and Spike Train

Chang Yunseo

2023.11.14

8-1 Entropy and Mutual Information

How much does the neural response tell us about the stimulus?

- Quantitatively
- What forms of Neural response are optimal

Information Theory

- Information Theory: Quantifying the ability of a coding scheme or a communication channel to convey information (stochastic & noisy process)
- Entropy: a measure of the theoretical capacity of a code to convey information
- Mutual information: how much of that capacity is actually used when the code is employed to describe a particular set of data

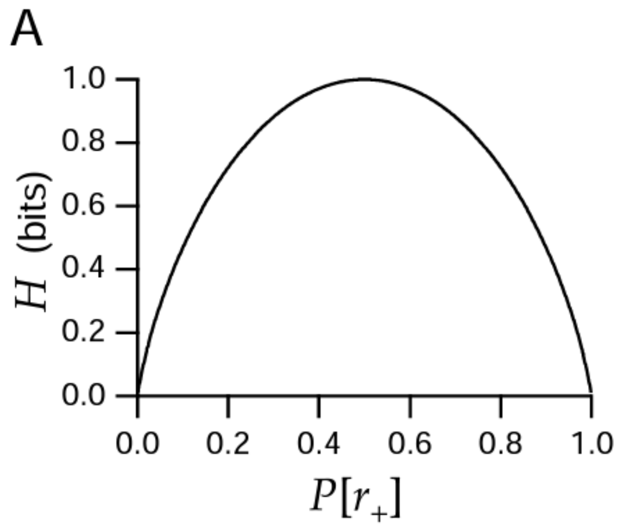
- Symbol: neuronal response / data: stimulus
- Simplified descriptions of the response of a neuron that reduce the number of possible symbols that need to be considered

Entropy

- Large range of different responds → interesting (irregular / unpredictable)
- observing a response spike-count rate r with possibility $P[r]$
- Entropy → surprise: $h(P[r]) = -log_2(P[r])$

: 1) decrease function. 2) $h(P[r_1]P[r_2]) = h(P[r_1]) + h(P[r_2])$ 3) information bits

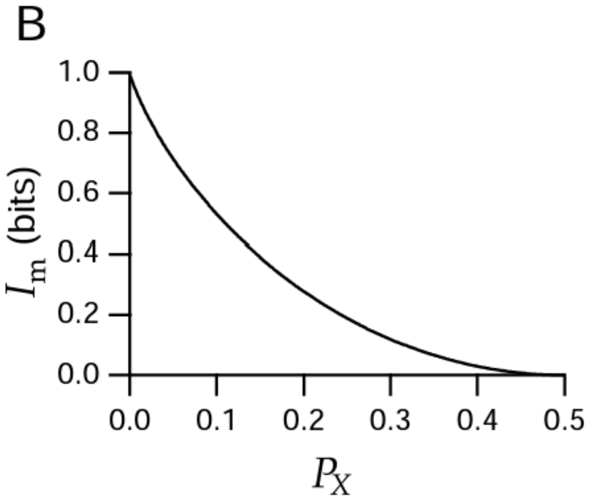
- Total Entropy $H = -\sum P[r]\log_2(P[r])$
- Same rate → P[r] = 0 or 1
- Have Two possible rate → $P[r_1] = P[r_2] = \frac{1}{2}$



Mutual information

What we can measure

- Different stimuli → Neural response different (does it correlate?)
- Mutual info: total response entropy – average response entropy on trials involving different stimulus
- $H_s = -\sum P[r|s] \log_2 P[r|s]$, $H_{noise} = \sum H_s P[s]$
- $I_m = \sum P[r,s] \log_2 \frac{P[r,s]}{P[r]P[s]}$ symmetric between r,s
- $\log_2 P[s|r]$: reduce total stimulus entropy
- $I_m = 1 + (1 - P_X) \log_2(1 - P_X) + P_X \log_2 P_X$



Mutual information

- Kullback-Leibler divergence
- $D_{KL}(P,Q) = \sum P[r] \log_2 \frac{P[r]}{Q[r]}$
- Normally associated with a distance measure, $D_{KL} \geq 0, D_{KL} = 0\ only\ at\ P = Q$
- Kullback-Leibler divergence between $P[r,s]$ $P[r]P[s]$

Continuous variables

- $H = -\sum p[r]\Delta r log_2(p[r]\Delta r) = -\sum p[r]\Delta r log_2(p[r]) - \log_2 \Delta r$
- $\Delta r \to 0, H \to \infty$: continuous variable measured with perfect accuracy ∞ entropy
- $\lim_{\Delta r \to 0}(H + \log_2 \Delta r) = -\int dr\ \ p[r] \log_2 p[r]$ Δr: limit of resolution
- $\lim_{\Delta r \to 0}(H_{noise} + \log_2 \Delta r) = \int ds \int dr\, p[s]\, p[r|s] \log_2 p[r|s]$
- $I_m = \int ds \int dr\, p[s]\, p[r|s] \log_2 \frac{p[r|s]}{p[r]}$

8-2 Information and Entropy Maximization

Entropy maximization for a Single neuron

- Maximum firing rate r_m
- $\int_0^{r_m} dr\, p[r] = 1$, maximize $-\int_0^{r_m} dr\ \ p[r] \log_2 p[r]$ → Lagrange multiplier
- $p[r] = \frac{1}{r_m}$ → $\mathrm{H} = \log_2 \frac{r_m}{\Delta r}$ Let $r = f(s)$
- $p[r]|\Delta r| = \frac{|f(s+\Delta s)-f(s)|}{r_m} = p[s]\Delta s$, $\frac{df}{ds} = r_m p[s]$
- $f(s) = r_m \int_s^{s_m} ds' p[s']$

Populations of Neurons

- Use vector $\vec{r} = (r_1, \cdots, r_N)$
- $H = -\int d\vec{r}\, p[\vec{r}] \log_2 p[\vec{r}] - Nlog_2 \Delta r$
- Consider $p[r_a] = \int \prod_{b \neq a} dr_b\, p[\vec{r}]$
- $H_a = -\int d\vec{r}\, p[\vec{r}]\ \log_2 p[r_a] - \log_2 \Delta r$, $H \leq \sum_a H_a$
- $\sum_a H_a - H = \int d\vec{r}\, p[\vec{r}]\ \log_2 \frac{p[\vec{r}]}{\prod_a p\,[r_a]}$: KL divergence

Populations of Neurons

- Entropy difference → redundancy
- To achieve Maximum population-response entropy..

1) Individual neurons must response independently
2) Have response probabilities that are optimized for whatever constraints are imposed

- $p[r_a]\ \ identical$

Populations of Neurons

- Covariance matrix : $Q_{ab} = \int d\vec{r}\, p[\vec{r}](r_a - \langle r \rangle)(r_b - \langle r \rangle) = \sigma_r^2 \delta_{ab}$
- Fix the covariance matrix, maximizes the entropy only if the statistics of the responses are gaussian

Application to Retinal Ganglion Cell Receptive Field

- Receptive field in Retina, LGN, primary visual cortex
- Maximize the amount of information that the associated neural responds convey about natural visual scenes in the presence of noise
- Only represent neural responds

Application to Retinal Ganglion Cell Receptive Field

- $L(t) = \int_0^\infty d\tau \int d\vec{x}\, D(\vec{x},\tau)s(\vec{x},t-\tau)$: linear estimation of the response of visual neuron.
- Contrast function $s(\vec{x},t)$, space time receptive field $D(\vec{x},\tau) = D_s(\vec{x})D_t(\tau)$
- $L_s = \int d\vec{x}\, D_s(\vec{x})s_s(\vec{x})$ $L_t(t) = \int_0^\infty d\tau\; D_t(\tau)s_t(t-\tau)$
- D: information carrying capacity
- All locations and directions are equivalent → same spatial structure

Application to Retinal Ganglion Cell Receptive Field

Centered at $\vec{a}$

- $L(\vec{a}) = \int d\vec{x}\ D_s(\vec{x} - \vec{a}) s_s(\vec{x})$
- We proceed as if there were a neuron corresponding to every continuous value of $\vec{a}$. This allows us to treat L($\vec{a}$) as a function of $\vec{a}$ and to replace sums over neurons with integrals over $\vec{a}$.

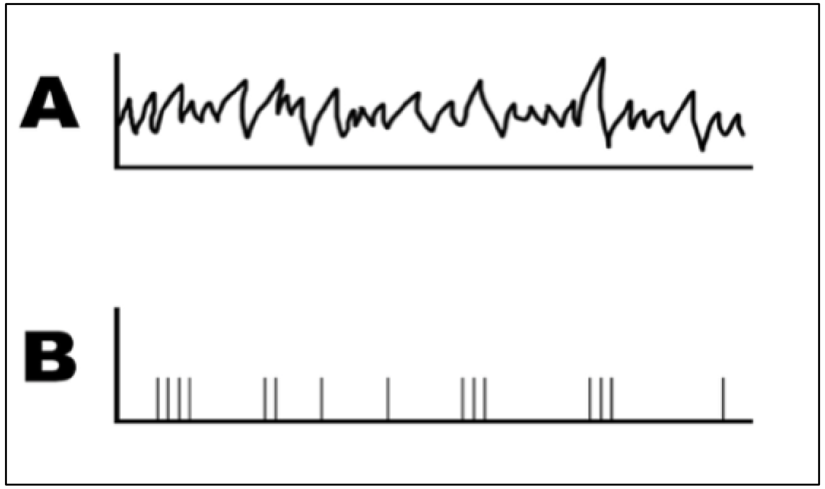
Spike Train and Poisson distribution

Spike train

- A sequence of recorded times at which a neuron fires an action potential
- 100mV over 1~2ms $\rightarrow$ each time can be considered by a single point

$$FR = \frac{\text{number of spikes}}{\Delta t}$$

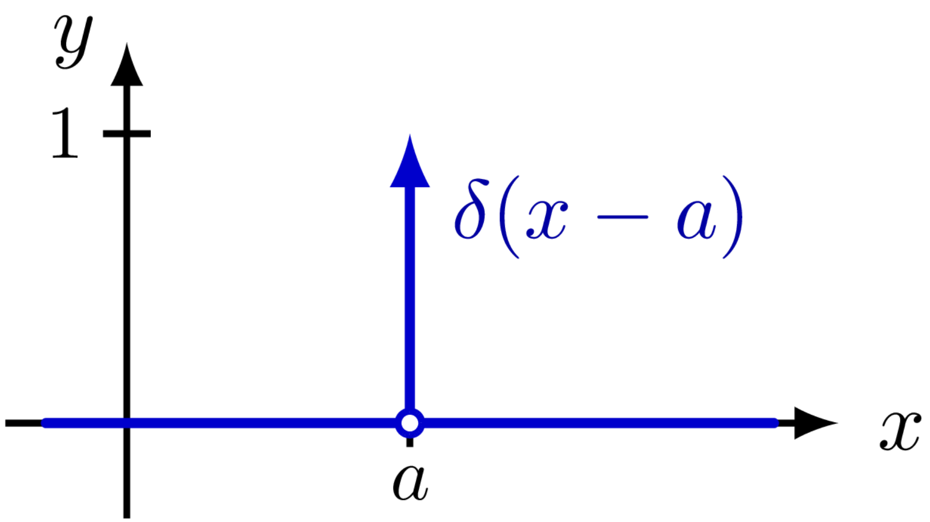
$$FR(t|x_t) = \lim_{\Delta t \to 0} \frac{P(\text{spike in } (t, t + \Delta t)|x_t)}{\Delta t}.$$



Example of a spike train. Graph A shows the recorded stimulus and graph B shows the recorded actions potentials during the stimulus.

Spike train

- Delta function



$$\int_{-\varepsilon}^{\varepsilon} \delta(x)dx = 1$$

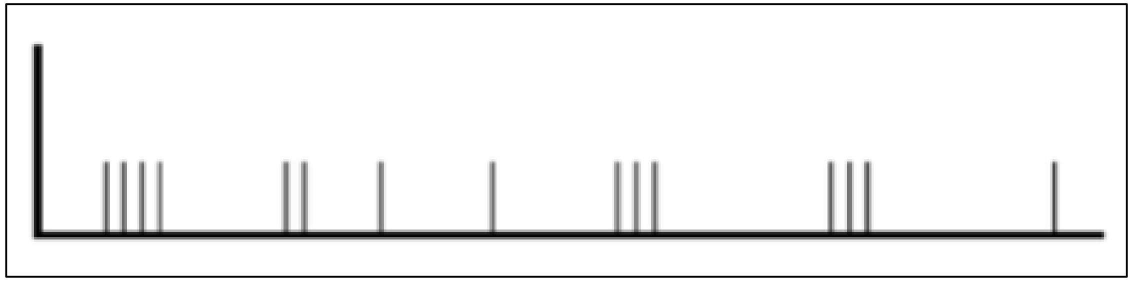
$$S(t) = \sum_k \delta(t - t^k)$$

Average number of spike trains per time

$$r = \langle S(t) \rangle = \lim_{T \to +\infty} \frac{1}{T} \int_0^T S(t)dt$$

Poisson Distribution

- discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time or space if these events occur with a known constant mean rate and independently of the time since the last event.



$$f(k;\lambda) = \Pr(X{=}k) = \frac{\lambda^k e^{-\lambda}}{k!},$$

$$\lambda = \operatorname{E}(X) = \operatorname{Var}(X).$$

$$P(k \text{ events in interval } t) = \frac{(rt)^k e^{-rt}}{k!}.$$

Average rate: r

Example

버스가 랜덤 하게 도착한다고 하자.

1시간 동안 도착하는 버스의 평균 도착 대수가 λ라면

1시간 동안 k개의 버스가 도착할 확률은 어떻게 되는가?

→ Poisson distribution

Derivation of Poisson distribution

사건이 일어날 확률은 동일하다고 하자. 그러면 이항분포로 나타낼 수 있음.

$$B(n,p,r) = \binom{n}{r} p^r (1-p)^{n-r}$$

이때 $p = \lambda/n$ 이고 $n \to \infty$

$$\lim_{n\to\infty} \frac{n!}{(n-r)!\,r!} \left(\frac{\lambda}{n}\right)^r \left(1-\frac{\lambda}{n}\right)^{n-r} = \frac{\lambda^r e^{-\lambda}}{r!}$$

Entropy and Information for Spike train

Entropy rate

$p[r]$: action potential 나타나는 rate r 일 확률
$< r > T$: action potential이 나타난 개수

- Firing rate 만으로는 spike train 모두 설명하기 어려움. → entropy 도입
- Entropy는 측정 시간이 증가하면 이에 비례하여 증가. → 단위 시간 당 entropy 값 생각 ($\dot{H}$)

$$H = -\langle r \rangle T \int_0^\infty d\tau \, p[\tau] \, \log_2(p[\tau] \, \Delta\tau)$$

$$\dot{H} \leq -\langle r \rangle \int_0^\infty d\tau \, p[\tau] \log_2(p[\tau] \Delta\tau) \, .$$

위 식은 뉴런끼리 independent할 때만 성립.
뉴런끼리 dependent하면 감소하므로 부등호 성립!

Entropy rate

$$\dot{H} \leq -\langle r \rangle \int_0^\infty d\tau\, p[\tau] \log_2(p[\tau]\Delta\tau)\,.$$

↓

$$\dot{H} = \frac{\langle r \rangle}{\ln(2)} \left(1 - \ln(\langle r \rangle \Delta\tau)\right)\,.$$

$$f(k;\lambda) = \Pr(X{=}k) = \frac{\lambda^k e^{-\lambda}}{k!},$$

Entropy rate

spike sequences of duration $T_s \ll T$ 도입

T_s continuous variable이지만 resolution Δt 생각 → discrete

B(t): $T_s/\Delta t$ bit binary number

$$\dot{H} = -\frac{1}{T_s}\sum_B P[B]\log_2 P[B] \,,$$ < 이에 의한 엔트로피

$B(t+T_s)$ & $B(t)$: correlate

Entropy rate

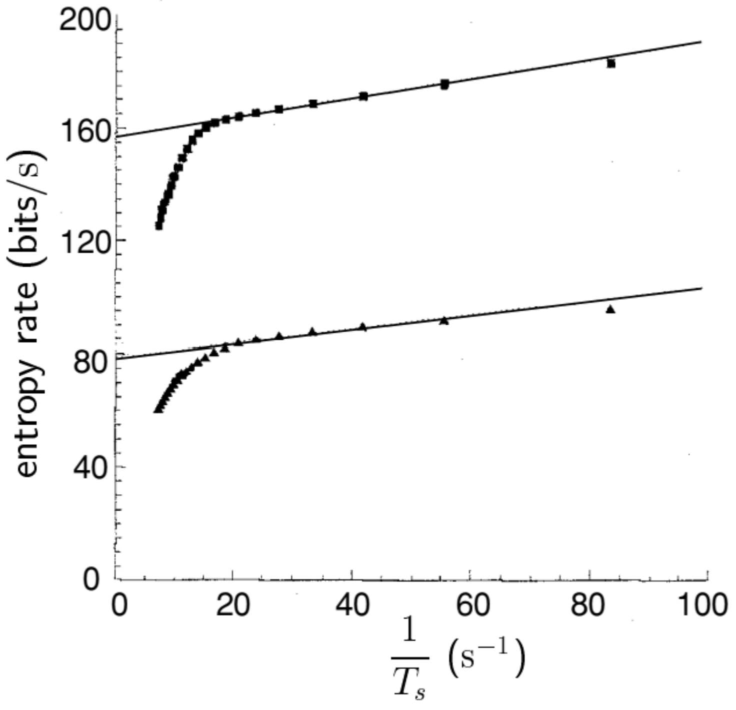
$B(t + T_s)$ & $B(t)$ correlation reduce the total-spike train entropy

T_s too small → $B(t + T_s)$ & $B(t)$ correlate

적당한 T_s 크기 존재

$T_s \rightarrow \infty$라면 true entropy can be measured

$\frac{1}{T_s} = 0$일 때 만나는 점 측정.



Entropy rate

Mutual information

$$\dot{H}_{\text{noise}} = -\frac{\Delta t}{T}\sum_t\left(\frac{1}{T_s}\sum_B P[B(t)]\log_2 P[B(t)]\right)$$

여기서 $\frac{\Delta t}{T}$ is the number of different t values being summed.

이전과 마찬가지로 $\frac{1}{T_s} = 0$일 때 만나는 점 측정하여 계산

Mutual information에서는 Δt 상쇄되지만 여기서는 여전히 엔트로피에 영향 미침.

Summary

Shannonís information theory can be used to determine how much a neural response tells both us and, presumably, the animal in which the neuron lives, about a stimulus. Entropy is a measure of the uncertainty or surprise associated with a stochastic variable, such as a stimulus. Mutual information quantifies the reduction in uncertainty associated with the observation of another variable, such as a response. The mutual information is related to the Kullback-Leibler divergence between two probability distributions. We defined the response and noise entropies for probability distributions of discrete and continuous firing rates, and considered how the information transmitted about a set of stimuli might be optimized.

Finally, we discussed how the information conveyed about dynamic stimuli by spike sequences can be estimated.