Solution to Homework Three

Cong Li(PB21081606)

1. Analyze Motion Detection Neuron in Fly

a) *Calculate and plot the spike-triggered average*

According to ref.1, $C(\tau) =< \frac{1}{n}\sum_{i=1}^{n} s(t_i - \tau) >$. So I sum the stimulus ocurring τ before all the spikes and average them to get the $C(\tau)$. And the result is shown in Fig.1.

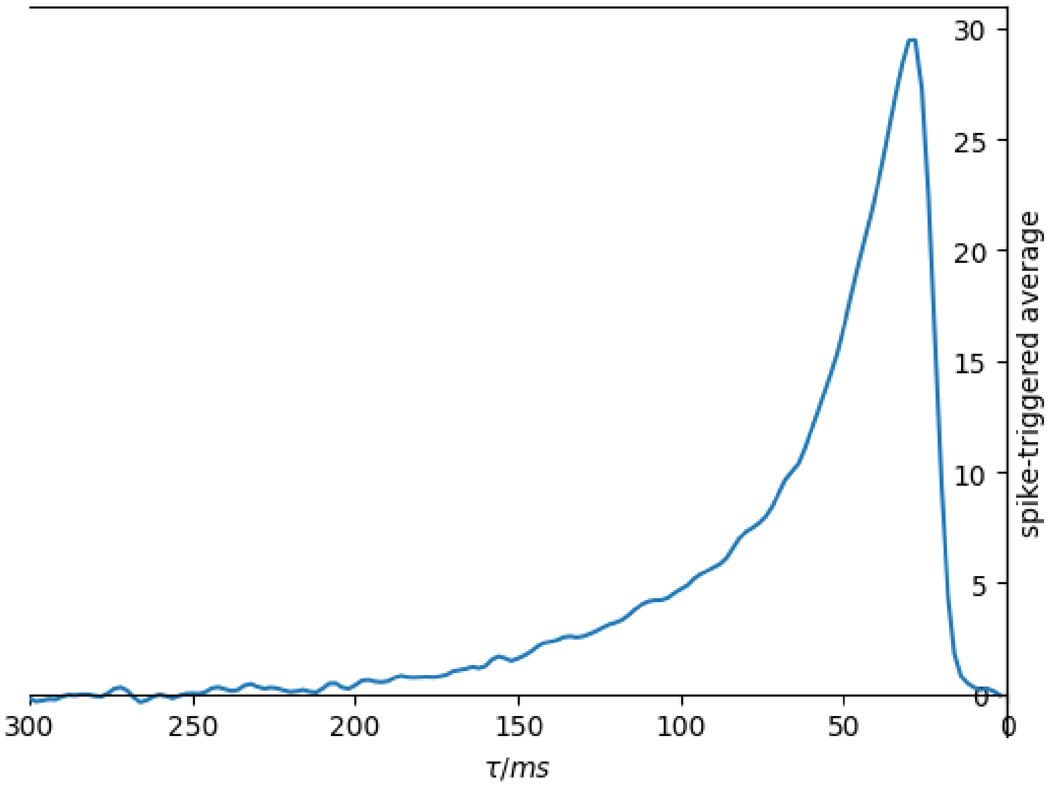

Figura 1. spike-triggeered average

As is shown in Fig.1, it's not the stimula happening closest to the spike that have the most contribution to that spike, but the stimula happening second closest to the spike. And those stimula happening far before the spike contribute little to the spike occurance.

b) *Construct a linear kernel to estimate the firing rate*

As we learned in class, the best linear kernel for white noise is $D(\tau) = \frac{<r>C(\tau)}{\sigma_s^2}$.

Because the mean value of stimulus is 0, I averaged the squared stimulus over the 20 minutes and multipied it by 0.002 seconds to get the $\sigma_s^2 = 5.107s$. And by dividing the total spike number happening in 300ms after specific time point t by 0.3 seconds, I got the firing rate of all time points. Averaging them, I had $< r >= 44.668Hz$.

In class, we've learned that $r_{est}(t) = r_0 + \int_0^\infty D(\tau)s(t-\tau)d\tau$, its discrete form is $r_{est}(t) = r_0 + \sum D(\tau)s(t-\tau)\Delta\tau$.

Here, $\Delta\tau = 2ms$. And it's reasonable to set $r_0 =< r >$. So by convolving stimulus $s(t)$ with the linear kernel we just got $D(\tau)$, I got the estimated firing rate and plotted it in Fig.2.

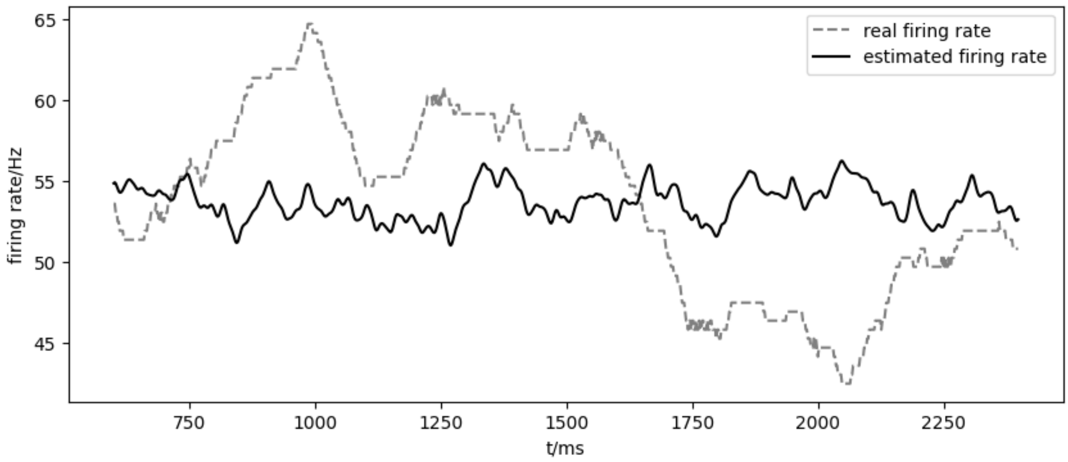

Figura 2. Comparsion between actual firing rate and estimated firing rate(interval=1800ms)

It seems that the linear kernel can roughly predict the trend of the firing rate but cannot reproduct the amplitude of the actual firing rate.

But I still have some doubts about this result. Because here I use the division of total spike number by time interval of 1800ms to calculate the real firing rate, in order to have a nice simulation result. However, if I use time interval of 300ms to calculate the real firing rate. The amplitude of the real firing rate would be significantly bigger than the estimated firing rate(Fig.3). It looks like the linear kernel is more suitable to predict the firing rate of a longer interval.

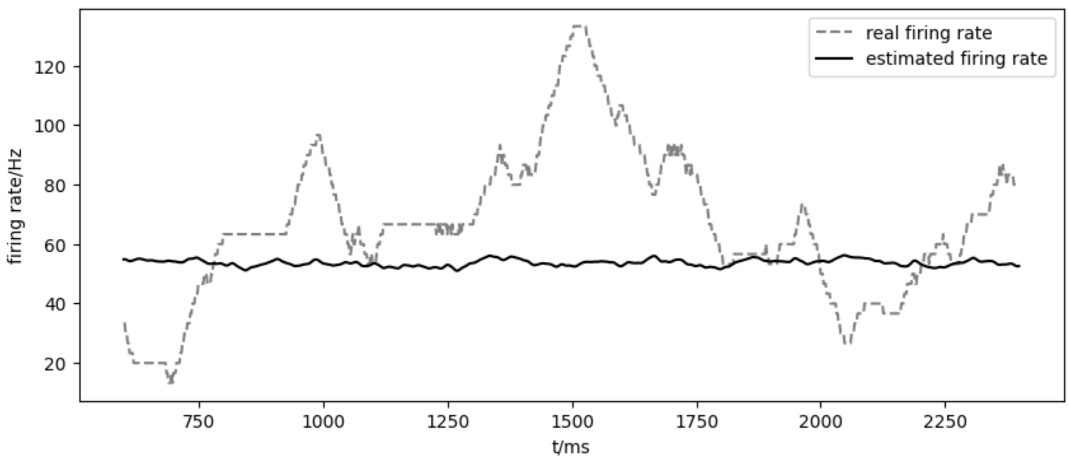

Figura 3. Comparsion between actual firing rate and estimated firing rate(interval=300ms)

c) *Use a Poisson generato to generate a synthetic spike train*

The average firing rate of my estimation is $44.536Hz$. So the estimation of simulated neuron firing times in 20 minutes is $44.536 \times 20 \times 60 = 53443$.

Because we know for a Possion process, the next event happening time $t \sim Exp(\lambda)$, an easy and quick idea to simulate a Possion process is to produce the interval for two spikes until the total time is bigger than 20 minutes.

Finally, I got the following result(Fig.4).

In Fig.4, the actual spike train and the simulated spike train are similar in that spikes would cluster

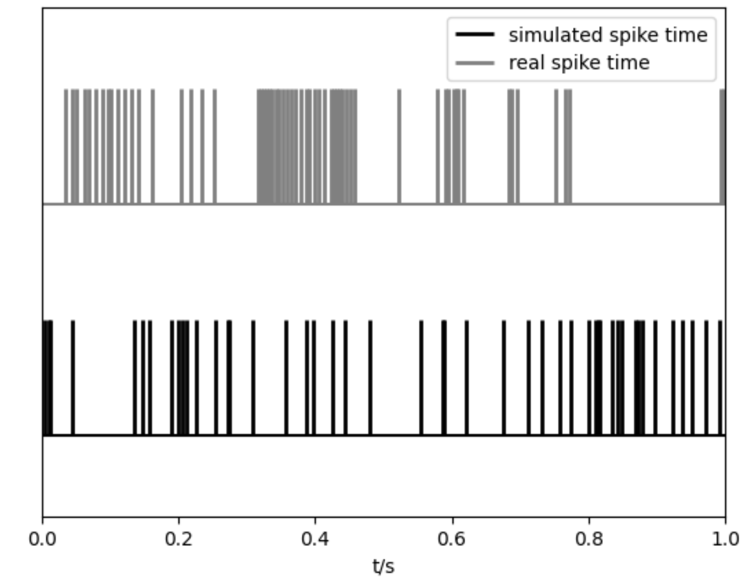

Figura 4. Comparsion between actual spike train and simulated spike train

sometimes. But it looks like the simulated spike train is sparser than the real spike train.

d) *Plot the autocorrelation function of the actual and the synthetic spike trains*

To plot the autocorrelation function of a spike train, I first chose a spike train running from 0 to 1s because one second is significantly greater than 2 milliseconds and therefore different time intreval of one second have similar identity. Then I calculated the firing rate in different time bins of 2 milliseconds. Finally I used the corerlate function in the numpy.

And here is the result(Fig.5).

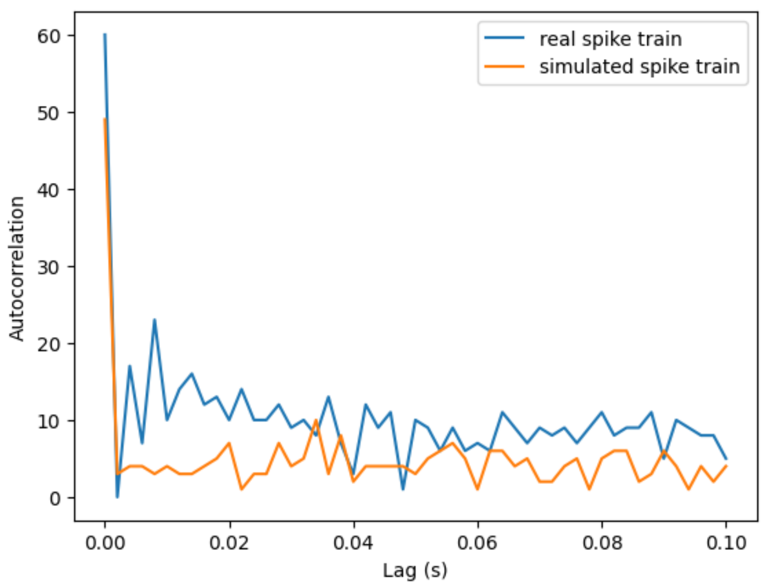

Figura 5. autocorrelation of actual and synthetic spike trains

We would notice that there a dip at a lag of 2 ms in the autocorrelation of the actual spike train but not the synthetic train. This is because of the existence of the refractory period of about 2 ms in actual neuron which causes that if it has fired in the a bin, the neuron cannot fire in the next bin.

Also, I want to share another autocorrelation plot result over the range of -1.00s to 1.00s. I think it looks very cool(Fig.6)!

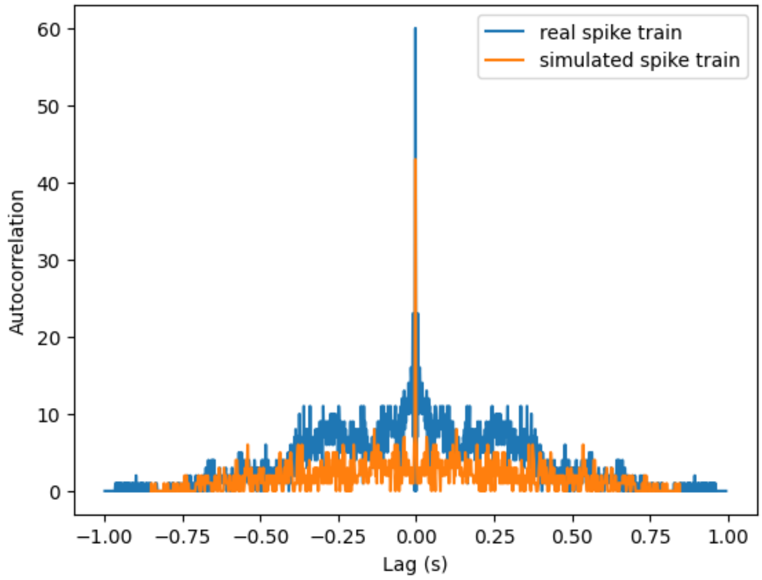

Figura 6. autocorrelation of actual and synthetic spike trains

And the following is the interspike interval histogram for both spike train(Fig.7).

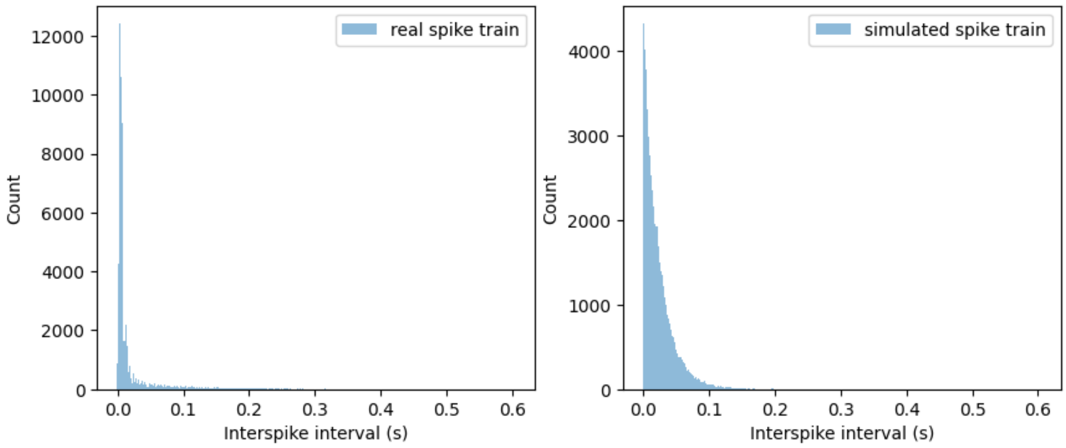

Figura 7. interspike interval histogram for actual and simulated spike train

Look clearly, we would find a dip below 6 ms in the histogram in the actual spike train but not for the simulated spike train. This is because actual neurons have refractory period of about 2 ms during which it is difficult for the neuron to fire again.

I'm sorry I cannot find any definition of the "coefficents of variation"mentioned by the teacher in all the lecture note. So I searched its meaning online. It seems that coefficent of variation of a spike train is the division of standard deviation by the mean value. Both of these parameters should come from the interspike interval. So I got the result shown in Table.1.

We have the coefficent of variation(CV) of the simualted spike train close to 1, which is understandable because the standard variation and mean value is the same for a expotential distribution. However, the CV of the real spike train is almost twice as much as of the simulated spike train. This is a symbol of higher variation in the real spike train which may be caused by the refractory period, adpatation etc.

	real spike train	**simulated spike train**
coefficent of variation	2.00855	1.00568

Tabela 1. coefficents of variation for two spike trains

2. Maximazatin of Entropy under Constraints

a) *X only takes positive values and its maximum is fixed*

Now the constraint is only the normalization requirement, in which case the problem can be described by the following formula.

$$\begin{aligned} &\max_{p(x)} -\int_0^{X_m} p(x)lnp(x)dx \\ &s.t. \int_0^{X_m} p(x)dx = 1 \end{aligned} \tag{1}$$

And X_m is the maximum of random variable X.

We can solve it by Lagrange multipier method.

$$L(p,\lambda) = -\int_0^{X_m} p(x)lnp(x)dx + \lambda(\int_0^{X_m} p(x)dx - 1) \tag{2}$$

Set $\frac{\delta L(p,\lambda)}{\delta p} = 0$, we will get

$$-(lnp(x)+1)+\lambda = 0 \Rightarrow p(x) = e^{\lambda-1} \tag{3}$$

Take $p(x) = e^{\lambda-1}$ into Eq.1, we will have $p(x) = \frac{1}{X_m}$, which obeys the uniform distribution.

b) *X only takes positive values and its mean value is fixed*

In this case, the problem can be described by the following equations

$$\begin{aligned} &\max_{p(x)} -\int_0^{X_m} p(x)lnp(x)dx \\ &s.t. \int_0^{\infty} p(x)dx = 1 \\ &\int_0^{\infty} p(x)\times xdx = C \end{aligned} \tag{4}$$

And C is a constant value.

Again, solve it by Lagrange multipier method.

$$L(p,\lambda,\mu) = -\int_0^{\infty} p(x)lnp(x)dx + \lambda(\int_0^{\infty} p(x)\times xdx - C) + \mu(\int_0^{\infty} p(x)dx - 1) \tag{5}$$

Set $\frac{\delta L(p,\lambda)}{\delta p} = 0$, we will get

$$-(lnp(x)+1)+\lambda x + \mu = 0 \Rightarrow p(x) = e^{\lambda x+\mu-1} \tag{6}$$

Take $p(x) = e^{\lambda x+\mu-1}$ into Eq.4, we will have $p(x) = \frac{1}{C}e^{-x/C}$, which obeys the expotential distribution.

c) *There is no constant on the range of X but its variance is given*

In this case, the problem can be described by the following equations

$$\begin{aligned} &\max_{p(x)} -\int_0^{X_m} p(x)lnp(x)dx \\ &s.t. \int_0^{\infty} p(x)dx = 1 \\ &\int_0^{\infty} p(x) \times (x-E(X))^2 dx = C \end{aligned} \tag{7}$$

And $E(X)$ is a mean value of X.

Again, solve it by Lagrange multipier method.

$$L(p,\lambda,\mu) = -\int_0^{\infty} p(x)lnp(x)dx + \lambda(\int_0^{\infty} p(x) \times (x-E(X))^2 dx - C) + \mu(\int_0^{\infty} p(x)dx - 1) \tag{8}$$

Set $\frac{\delta L(p,\lambda)}{\delta p} = 0$, we will get

$$-(lnp(x)+1) + \lambda(x-E(X))^2 + \mu = 0 \Rightarrow p(x) = e^{\lambda(x-E(X))^2+\mu-1} \tag{9}$$

Take $p(x) = e^{\lambda(x-E(X))^2+\mu-1}$ into Eq.7, we will have $p(x) = \frac{1}{\sqrt{2\pi C}}e^{-(x-E(X))^2/2C}$, which obeys the normal distribution.

d) *X is an N-dimensional continuous random variable with constraint on the total variance*

The value $\vec{X}$ can take will be expressed by $(x_1, x_2, ..., x_N)$. As the teacher expressed the total variance as $\sum_i < x_i^2 >$, I think there is a hidden condition that $E(x_i) = 0$. Now the problem can be described as

$$\begin{aligned} &\max_{p} -\int_{R^N} p(x_1,x_2,...,x_N)ln(p(x_1,x_2,...,x_N))dx_1dx_2...dx_N \\ &s.t. \int_{R^N} p(x_1,x_2,...,x_N)dx_1dx_2...dx_N = 1 \\ &\int_{R^N} (x_1^2+x_2^2+...+x_N^2)p(x_1,x_2,...,x_N)dx_1dx_2...dx_N = N\sigma^2 \end{aligned} \tag{10}$$

So I constructed such a Lagrange function

$$\begin{aligned} L(p,\lambda,\mu) = &-\int_{R^N} p(x_1,x_2,...,x_N)ln(p(x_1,x_2,...,x_N))dx_1dx_2...dx_N + \\ &\lambda[\int_{R^N} (x_1^2+x_2^2+...+x_N^2)p(x_1,x_2,...,x_N)dx_1dx_2...dx_N - N\sigma^2] + \\ &\mu(\int_{R^N} p(x_1,x_2,...,x_N)dx_1dx_2...dx_N - 1) \end{aligned} \tag{11}$$

Now set $\frac{\delta L}{\delta p} = 0$, we will get

$$-(lnp(\vec{X})+1) + \lambda\sum_i x_i^2 + \mu = 0 \Rightarrow p(\vec{X}) = e^{\lambda\sum_i x^2+\mu-1} \tag{12}$$

Take $p(\vec{X}) = e^{\lambda\sum_i x^2+\mu-1}$ into Eq.7, we will have $p(\vec{X}) = (\frac{\pi}{2\sigma^2})^{N/2}e^{-\frac{\sum_i x_i^2}{2\sigma^2}}$, which obeys the multi-dimensional Gaussion distribution.

Note: we also need to prove that the extreme value solution gotten by the Lagrange multipier method is the maximal value solution. And we would notice that $\frac{\delta^2 L}{\delta p^2} < 0$ and we only get one extreme value solution which means that the extreme value solution is absolutely the maximal value solution.

e) *Prove the entropy calculation formula of the multivariate Gaussian distribution*

I'm sorry that owing to my poor understanding of the linear algebra, I had difficultiy solving this problem. So I tried to search the demonstration process online and here is my learning result(I mean that I have tried to better the demonstration postered by others).

$$\begin{aligned}
H[\vec{X}] &= -\int_{R^N} N(\vec{X}|\vec{\mu},\sum)ln(N(\vec{X}|\vec{\mu},\sum))d\vec{X} \\
&= -E[ln(N(\vec{X}|\vec{\mu},\sum))] \\
&= -E[ln((2\pi)^{-D/2})|\sum|^{-1/2}exp(-\frac{1}{2}(\vec{x}-\vec{\mu})^T\sum{}^{-1}(\vec{x}-\vec{\mu}))] \quad \text{definition of multivariable Gaussion} \\
&= \frac{D}{2}ln(2\pi)+\frac{1}{2}|\sum|+\frac{1}{2}E[(\vec{x}-\vec{\mu})^T\sum{}^{-1}(\vec{x}-\vec{\mu})]
\end{aligned} \tag{13}$$

Consider the third term in the equation above

$$\begin{aligned}
E[(\vec{x}-\vec{\mu})^T\sum{}^{-1}(\vec{x}-\vec{\mu})] &= E[tr((\vec{x}-\vec{\mu})^T\sum{}^{-1}(\vec{x}-\vec{\mu}))] \\
&= E[tr(\sum{}^{-1}(\vec{x}-\vec{\mu})(\vec{x}-\vec{\mu})^T)] \\
&= tr(E[\sum{}^{-1}(\vec{x}-\vec{\mu})(\vec{x}-\vec{\mu})^T]) \\
&= tr(\sum{}^{-1}E[(\vec{x}-\vec{\mu})(\vec{x}-\vec{\mu})^T]) \\
&= tr(\sum{}^{-1}\sum) \\
&= D
\end{aligned} \tag{14}$$

Therefore, combining Eq.13 and Eq.14, we get $H[\vec{X}] = \frac{1}{2}|\sum| + \frac{D}{2}(ln(2\pi)+1)$. ■

3. K-L divergence

Because $log(x)$ is a concave function, $log(1/x) = -log(x)$ is a convex function. According to Jensen's inequality, we have

$$D_{KL} = \sum_i P(x_i)\times(-log(\frac{Q(x_i)}{P(x_i)})) \geq -log(\sum_i P(x_i)\times\frac{Q(x_i)}{P(x_i)}) = -log(\sum_i Q(x_i)) = 0 \tag{15}$$

So, D_{KL} will be equal to 0 only when $Q(x)/P(x) = C$ and C is a constant value. However, because $\sum Q(x) = C\sum P(x) = C = 1$, $D_{KL} = 0$ will be taken only when $P(x) = Q(x)$.

4. Fisher Information and Mutual Information

a) *Compute the Fisher Information of the system*

As $r_i = \omega x + z_i$ and $z_i \sim N(0, \sigma_i^2)$, $x \sim N(0, \sigma_0^2)$, $r_i|s = x \sim N(\omega x, \sigma_i^2)$.
Because z_i are uncorrelated, $p(\vec{r}|s) = \prod_{i=1}^{N} p(r_i|s)$.
So

$$\begin{aligned} ln(p(\vec{r}|s)) &= ln(\prod_{i=1}^{N} p(r_i|s)) \\ &= ln(\prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_i} exp(-\frac{(r_i - \omega x)^2}{2\sigma_i^2})) \\ &= \sum_{i=1}^{N} [ln \frac{1}{\sqrt{2\pi}\sigma_i} - \frac{(r_i - \omega x)^2}{2\sigma_i^2}] \end{aligned} \tag{16}$$

We will get

$$\frac{\partial^2 ln p(\vec{r}|s)}{\partial s^2} = -\sum_{i=1}^{N} \frac{\omega^2}{\sigma_i^2} \tag{17}$$

According to the definition of Fisher information, we will get the Fisher information of this system

$$\begin{aligned} I_F &= < -\frac{\partial^2 ln(p(\vec{r}|s))}{\partial s^2} > \\ &= \int d\vec{r} p(\vec{r}|s)(-\frac{\partial^2 ln p(\vec{r}|s)}{\partial s^2}) \\ &= \int d\vec{r} p(\vec{r}|s) \sum_{i=1}^{N} \frac{\omega^2}{\sigma_i^2} \\ &= \sum_{i=1}^{N} \frac{\omega^2}{\sigma_i^2} \int d\vec{r} \prod_{i=1}^{N} p(r_i|s) \\ &= \sum_{i=1}^{N} \frac{\omega^2}{\sigma_i^2} \end{aligned} \tag{18}$$

b) *Find the ML estimator of* $\hat{x}$

To find the stimulus x to maximize $p(\vec{r}|s = x)$, we should find the x to maximize $ln(p(\vec{x}|s = x))$. As shown in the Eq.16, $ln(p(\vec{r}|s = x)) = \sum_{i=1}^{N} [ln \frac{1}{\sqrt{2\pi}\sigma_i} - \frac{(r_i - \omega x)^2}{2\sigma_i^2}]$.

Set $\frac{\partial ln(p(\vec{r}|s=x))}{\partial s} = 0$, we will get $x = \sum_{i=1}^{N} \frac{r_i}{\sigma_i^2} / (\omega \sum_{i=1}^{N} \frac{1}{\sigma_i^2})$. Because $\frac{\partial^2 ln(\vec{r}|s=x)}{\partial s^2} = -\sum_{i=1}^{N} \frac{\omega^2}{\sigma_i^2} < 0$, function $ln(p(\vec{r}|s = x))$ will take the maximum at the point $x = \sum_{i=1}^{N} \frac{r_i}{\sigma_i^2} / (\omega \sum_{i=1}^{N} \frac{1}{\sigma_i^2})$. That is to say, $\hat{x} = \sum_{i=1}^{N} \frac{r_i}{\sigma_i^2} / (\omega \sum_{i=1}^{N} \frac{1}{\sigma_i^2})$.

So

$$\begin{aligned} < (x - \hat{x})^2 > &= E[Var(\hat{x})|x] \\ &= \int_R p(x = k) \frac{\sum \frac{1}{\sigma_i^2}}{(\omega^2 \sum \frac{1}{\sigma_i^2})^2} dk \\ &= \frac{1}{\sum_{i=1}^{N} \frac{\omega^2}{\sigma_i^2}} \end{aligned} \tag{19}$$

According to Eq.18 and Eq.20, we find that $< (x - \hat{x})^2 >= \frac{1}{I_F}$. This result conforms the Cramér-Rao bound–$\sigma_{est}^2 \geq \frac{(1+b_{est})^2}{I_F}$(here $b_{est} = 0$).

c) *Find a Bayesian estimator* $\hat{x}$ *that would maximize the posterior distribution* $p(x|\vec{r})$

Finding the $\hat{x}$ to maximize the $p(x|\vec{r})$ is equivalent to finding the $\hat{x}$ to maximize $ln(p(x|\vec{r}))$ because $p(\vec{r}|x), p(x)$ and $p(\vec{r})$ is greater than 0.

According to Eq.17, $ln(p(\vec{r}|x)) = \sum_{i=1}^{N}[ln\frac{1}{\sqrt{2\pi}\sigma_i} - \frac{(r_i-\omega x)^2}{2\sigma_i^2}]$.

And owing to $x \sim N(0, \sigma_0^2)$, $ln(p(x)) = ln\frac{1}{\sqrt{2\pi}\sigma_0} - \frac{x^2}{2\sigma_0^2}$.

So

$$\begin{aligned} ln(p(x|\vec{r})) &= ln(p(\vec{r}|x)) + ln(p(x)) - ln(p(\vec{r})) \\ &= \sum_{i=1}^{N}[ln\frac{1}{\sqrt{2\pi}\sigma_i} - \frac{(r_i-\omega x)^2}{2\sigma_i^2}] + ln\frac{1}{\sqrt{2\pi}\sigma_0} - \frac{x^2}{2\sigma_0^2} - \\ &\sum_{i=1}^{N}[ln\frac{1}{\sqrt{2\pi(\omega^2\sigma_0^2+\sigma_i^2)}} - \frac{r_i^2}{2(\omega^2\sigma_0^2+\sigma_i^2)}] \end{aligned} \tag{20}$$

Set $\frac{\partial ln(p(x|\vec{r}))}{\partial x} = 0$, we will get $x = \sum_{i=1}^{N}\frac{\omega r_i}{\sigma_i^2}/(\sum_{i=1}^{N}\frac{\omega^2}{\sigma_i^2} + \frac{1}{\sigma_0^2})$. Because $\frac{\partial^2 ln(p(x|\vec{r}))}{\partial x^2} = -\sum_{i=1}^{N}\frac{\omega^2}{\sigma_i^2} - \frac{1}{\sigma_0^2} < 0$, function $ln(p(x|\vec{r}))$ will take the maximum at the point $x = \sum_{i=1}^{N}\frac{\omega r_i}{\sigma_i^2}/(\sum_{i=1}^{N}\frac{\omega^2}{\sigma_i^2} + \frac{1}{\sigma_0^2})$. That is to say, $\hat{x} = \sum_{i=1}^{N}\frac{\omega r_i}{\sigma_i^2}/(\sum_{i=1}^{N}\frac{\omega^2}{\sigma_i^2} + \frac{1}{\sigma_0^2})$.

So

$$\begin{aligned} <(x-\hat{x})^2> &= E[Var(\hat{x})|x] \\ &= \int_R p(x=k)\frac{\sum\frac{\omega^2}{\sigma_i^2}}{(\sum\frac{\omega^2}{\sigma_i^2}+\frac{1}{\sigma_0^2})^2}dk \\ &= \frac{\sum\frac{\omega^2}{\sigma_i^2}}{(\sum\frac{\omega^2}{\sigma_i^2}+\frac{1}{\sigma_0^2})^2} \end{aligned} \tag{21}$$

We also have

$$b_{est} = E[E(\hat{x}) - x|x] = \int_R p(x=k)\frac{\frac{x}{\sigma_0^2}}{(\sum\frac{\omega^2}{\sigma_i^2}+\frac{1}{\sigma_0^2})^2} = 0 \tag{22}$$

According to Eq.18 and Eq.21, we would find that $<(x-\hat{x})^2>> \frac{1}{I_F}$, because $<(x-\hat{x})^2>$ is greater than $\frac{\sum\frac{\omega^2}{\sigma_i^2}}{(\sum\frac{\omega^2}{\sigma_i^2})^2}$.

Based on the calculation in b) and c), we found that when $\sigma_0^2 \gg 1$, $\hat{x}_{Bayesian} \approx \hat{x}_{MSL}$. So if σ_0^2 is not enough greater than 1, the difference of these two estimators would be significant.

As for the dependence on σ_0^2 of $\bar{x}_{Bayesian}$, I think this is because σ_0^2 has an effect on $p(x)$. So those x close to 0 would have higher possibility to be included in the estimation.

d) *Compute the Mutual Information between the neuronal responses and the stimulus*

As $r_i = \omega x + z_i$ and $z_i \sim N(0, \sigma_i^2), x \sim N(0, \sigma_0^2)$, we have $r_i \sim N(\omega x, \sigma_i^2), r_i \sim N(0, \omega^2\sigma_0^2 + \sigma_i^2)$. Because $< z_i z_j >= \delta_{i,j}^2$, $\vec{r}$ obeys multi-dimensional Gaussian distribution and correlation is 0.

So according to Eqs.13-14,

$$H_{total} = \frac{1}{2}ln(\prod_{i=1}^{N}(\omega^2\sigma_0^2+\sigma_i^2)) + \frac{N}{2}(1+ln(2\pi)) \tag{23}$$

and

$$H_{noise} = \int_R p(x)H[\vec{r}|x]dx = \int_R p(x)(\frac{1}{2}ln(\prod_{i=1}^{N}\sigma^2) + \frac{N}{2}(1+ln(2\pi)))dx = \frac{1}{2}ln(\prod_{i=1}^{N}\sigma_i^2) + \frac{N}{2}(1+ln(2\pi)) \tag{24}$$

Therefore, mutual information between the neuronal responses and the stimulus is

$$\begin{aligned} I_m &= H_{total} - H_{noise} \\ &= \frac{1}{2}ln(\prod_{i=1}^{N}(\frac{\omega^2\sigma_0^2}{\sigma_i^2}+1)) \end{aligned} \tag{25}$$

If we assume all $\sigma_i = \sigma$, $I_F = \frac{N\omega^2}{\sigma^2}$, $I_m = \frac{N}{2}ln(\frac{\omega^2\sigma_0^2}{\sigma^2}+1)$. So $I_m = \frac{N}{2}ln(\frac{I_F\sigma_0^2}{N}+1)$.

5. Acknowledgement

The successful completion of the work would be impossible without the help of ChatGPT, the TA's patient instructions and the following blog post calculation of entropy of multivaiate Gaussian distribution.

6. Reference

[1] Dayan, P., & Abbott, L. F. (2001). Theoretical neuroscience: computational and mathematical modeling of neural systems.