Mathematical properties of Gaussian process neural code

Nischal Mainali

November 18, 2024

Contents

§1 Bckground

Theoretical neuroscience involves a growing list of mathematical tools and objects, and among them probability is an important player that appears widely. First, and perhaps most commonly, it appears in theoretical models as "noise" to mimic inherent randomness observed in neural recordings. One can then ask questions such as: what is the nature and source of this noise? and, how do neurons and neural population adapt to and function in face of such noise? Secondly, they are involved in stylized problems of neural learning where one would like to understand how various plasticity mechanisms can be used to learn by population of neurons when provided with stochastic learning data and/or stochastic learning rule. Thridly, random variables are commonly used to model synaptic weight strengths to explain observed neural dynamics in the brain and to model how such dynamics can perform computations. Fourthly, randomness is used instrumentally in models of computations, memories or dynamics to get an analytical handle on their properties and includes the large class of statistical (mean) field theories deployed in neuroscience. Here, we will be concerned with a slightly different use of randomness: modeling neuronal representations of external variables as random code generated by neural population and their statistical properties. In particular, we will be concerned with continuous external variable (such as space) represented by a particular continuous stochastic process: Gaussian process [1].

But before, we briefly recall discrete Gaussian random variables and give definition for a Gaussian process.

§1.1 Discrete Gaussian variable

Definition

A real valued random variable ξ is said to be Gaussian if it has the density function

$$\phi_\xi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{x-m}{2\sigma^2}\right) \tag{1}$$

for some $m \in \mathbb{R}$ and $\sigma \geq 0$.

Similarly, a vector $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_p)$ of p random variables is said to have a p-variate Gaussian (normal) distribution if every linear combination of its components $\mathbf{a} \cdot \xi = \sum_k a_k \xi_k$ has a normal distribution. The variables $\xi_1, \ldots, \xi_p$ are said to be jointly Gaussian with mean vector $\mathbf{m} = \langle \boldsymbol{\xi} \rangle$ and non-negative definite $p \times p$ covariance matrix

$$\boldsymbol{\Sigma} = C(\boldsymbol{\xi}; \boldsymbol{\xi}) = \left\langle (\boldsymbol{\xi} - \mathbf{m})^\top (\boldsymbol{\xi} - \mathbf{m}) \right\rangle$$

If the determinant of $\boldsymbol{\Sigma}$ is positive, the distribution of $\boldsymbol{\xi}$ is non-singular and has a density

$$\phi_{\boldsymbol{\xi}}(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}\sqrt{\det \boldsymbol{\Sigma}}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m})\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\mathbf{m})'}. \tag{2}$$

§1.1.1 Conditional Gaussian

The multivariate Gaussian distribution has the useful property that conditional on observations of a subset of variables, the unobserved variables are also Gaussians with the conditional expectation as a linear combination of the observations and variances and covariances that are independent of the observed values. Let $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)$ and $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_m)$ be two jointly Gaussian vectors with means:

$$\langle \boldsymbol{\xi} \rangle = m_{\boldsymbol{\xi}}, \quad \langle \boldsymbol{\eta} \rangle = m_{\boldsymbol{\eta}},$$

and with covariance matrix:

$$\boldsymbol{\Sigma} = C((\boldsymbol{\xi}, \boldsymbol{\eta}); (\boldsymbol{\xi}, \boldsymbol{\eta})) = \begin{pmatrix} \boldsymbol{\Sigma}_{\boldsymbol{\xi}\boldsymbol{\xi}} & \boldsymbol{\Sigma}_{\boldsymbol{\xi}\boldsymbol{\eta}} \\ \boldsymbol{\Sigma}_{\boldsymbol{\eta}\boldsymbol{\xi}} & \boldsymbol{\Sigma}_{\boldsymbol{\eta}\boldsymbol{\eta}} \end{pmatrix}.$$

Theorem 1.1 (Gaussian Conditioning)

The conditional distribution of $\boldsymbol{\xi}$ given $\boldsymbol{\eta} = \mathbf{y}$ is Gaussian with conditional mean matrix

$$\begin{aligned} \langle \boldsymbol{\xi} \mid \boldsymbol{\eta} = \mathbf{y} \rangle = \widehat{\boldsymbol{\xi}}(\mathbf{y}) &= \langle \boldsymbol{\xi} \rangle + C(\boldsymbol{\xi}; \boldsymbol{\eta}) \boldsymbol{\Sigma}_{\eta\eta}^{-1} (\mathbf{y} - \langle \boldsymbol{\eta} \rangle)^{\top} \\ &= m_{\boldsymbol{\xi}} + \boldsymbol{\Sigma}_{\xi\eta} \boldsymbol{\Sigma}_{\eta\eta}^{-1} (\mathbf{y} - m_{\boldsymbol{\eta}})^{\top}. \end{aligned} \tag{3}$$

The conditional covariance is

$$\boldsymbol{\Sigma}_{\boldsymbol{\xi}\boldsymbol{\xi}|\boldsymbol{\eta}} = \left\langle \left((\boldsymbol{\xi} - \widehat{\boldsymbol{\xi}}(\boldsymbol{\eta}))^{\top} (\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}(\boldsymbol{\eta})) \right) \right\rangle = \boldsymbol{\Sigma}_{\boldsymbol{\xi}\boldsymbol{\xi}} - \boldsymbol{\Sigma}_{\boldsymbol{\xi}\eta} \boldsymbol{\Sigma}_{\eta\eta}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{\eta}\boldsymbol{\xi}}. \tag{4}$$

§1.2 Gaussian process

Definition 1.2

A stochastic process $\{f(x), x \in \mathbb{R}\}$ is a Gaussian process if every linear combination $S = \sum_k a_k f(x_k)$ for real a_k and $x_k \in \mathbb{R}$ has a Gaussian distribution.

Equivalently, $f(x)$ is a Gaussian process if for every finite discrete k points it's evaluated at $(f(x_1), \cdots, f(x_k))$ are distributed as multivariate Gaussians.

Similar to how multivariate Gaussian is defined using one index mean vector and two index covariance matrix, Gaussian process is defined using a one-variable mean function When we went from univariate to multivariate Gaussian we went from single index variable f to finite discrete indexed variable $f = (f_1, \ldots, f_L)$. Gaussian process generalize this process further and move to a continuous indexed random variable $f(x)$, which is then a **function** over the index $\mathbb{R}$.

$$m(x) = \langle f(x) \rangle,$$

and two-variable covariance function

$$r(x, y) = \langle (f(x) - m(x)) (f(y) - m(y)) \rangle.$$

§1.3 Origins of random code

Consider a population $i = 1 \dots N$ of spatially selective neurons presynpatic to a CA1 place cell with bounded spatial response function $u_i(x)$ for $x \in [0, L]$. We consider first, for simplicity, a scenario in which all input neurons have the same tuning curve up to translation, with preferred firing locations x_i that uniformly and densely tiles the space:

$$u_i(x) = u_0(x - x_i)$$

The input to the CA1 cell, $f(x)$ is a weighted sum of presynpatic activity:

$$f(x) = \sum_{i=1}^{N} W_i u_i(x),$$

where we assume that the weights W_i are drawn independently from a distribution with zero mean, and a finite second moment. Using the central limit theorem for multidimensional stimuli [2, 3], we get that for any discrete set of locations x_k, $f(x_i)$ are jointly distributed as multivariate normal in the large N limit. Hence, $f(x)$ is a Gaussian process in this limit. The covariance function of the Gaussian process must be transitionally invariant, due to our assumption of uniform tiling of inputs in the input layer. Indeed,

$$\begin{aligned}
\langle f(x)f(y)\rangle &= \sum_{i,j=1}^{N} \langle W_i W_j\rangle\, u_i(x)u_i(y) = \mathrm{Var}\,(w) \sum_{i=1}^{N} u_i(x)u_i(y) \\
&\approx \mathrm{Var}\,(w)\,\rho \int_L u_0(x)u_0(x+\Delta x)\,\mathrm{d}x\,; \quad \rho = \frac{N}{L} \\
&\text{Assume for simplicity: } \mathrm{Var}(w) = \frac{L}{N} \\
&= r(y-x) = r(\Delta x)
\end{aligned}$$

In particular, $r(0) = r_0$ is the variance of the process.

Example

Suppose individual tuning curves are shaped as Gaussian firing fields, $r(\Delta x)$ is a Gaussian with a width equal to twice the width of the individual firing fields.

$$\begin{aligned}
u_i(x) = Re^{-\frac{(x-c_j)^2}{2\sigma^2}} \implies r(\Delta x) &\approx R^2 \int_0^1 dc_j \exp\left(-\frac{(x-c_j)^2 + (y-c_j)^2}{2\sigma^2}\right) \\
&\approx R^2\sqrt{\pi\sigma^2}\exp\left(-\frac{\Delta x^2}{4\sigma^2}\right)
\end{aligned}$$

In our model, the underlying index $x \in [0, L]$ is provided by the environment and represents the location of the animal. The Gaussian process, then, represents the subthreshold activity of a cell in CA1 in hippocampus and is composed to sum of many synaptic activation at each point in the space, which can correspond to a Gaussian process in limit of large input. These subthreshold activation are filtered through cellular thresholding (and possibly incident inhibitory input to the cell) and when it crosses a threshold $\theta\sqrt{r_0}$ (thresholds should be in the order of variance of the process to be meaningful) resulting in place fields of the neurons.

§2 Structure of excursion of GP

In our setting, the inputs to a neuron is a sample from a Gaussian process. This input is then processed by the neuron to generate neural (firing rate) activity. This can be a highly non-linear process but always involves some sort of thresholding operation:

$$\Theta(h) = \max\left(0, f - \theta\sqrt{r_0}\right)$$

This gives rise to "excursion sets" of the process when it is above such threshold, which are region of non-zero firing rates in the backdrop of zero activity. Notice that various features of these excursion sets either don't depend on further non-linearity such as number or sizes of the excursion sets or depends on it in a trivial way such as height and derivatives. With this, the simplest question one can begin to ask is how many excursion sets we might get in in the space of size L. This is, of course, a random variable, but a famous formula found by Kac and Rice (Kac-Rice formula) tells us how to count the expected number of threshold crossing event of a stochastic process.

§2.1 Kac-Rice formula

First, we will state and show the counting formula, proved by Kac and Rice[4, 5], which is true for any functions or process under mild regularity conditions (C^1 functions). Then, we will apply it to Gaussian process to find the expected number of excursion sets.

Theorem 2.1 (Kac-Rice formula)

For every C^1 function $f : \mathbb{R} \to \mathbb{R}$, that is compact on $[a, b]$, and has no double roots, the number of threshold crossings $N_f(T)$ at the threshold T is given by

$$N_f(T) = \lim_{\epsilon \to 0} \int_a^b \eta_\epsilon(f(x) - T) \left|f'(x)\right| \mathrm{d}x \tag{5}$$

where

$$\eta_\epsilon(g) = \frac{1}{2\epsilon} \mathbb{1}\left\{|g| < \epsilon\right\}.$$

Firthermore, in the limit $\epsilon \to 0$:

$$N_f\left(T\right) = \int_L \delta\left(f(x) - T\right) \left|f'(x)\right| \mathrm{d}x\,, \tag{6}$$

where $\delta(.)$ is the delta function.

Proof. As $\epsilon \to 0$, the number of intervals such that $\{x \in [a,b] \mid |f(x) - T| < \epsilon\}$ is a union of intervals around the threshold crossings, and we choose a particular interval around a threshold crossing $[c,d]$, then:

$$\begin{aligned} N_f(T) &= \lim_{\epsilon \to 0} \int_a^b \eta_\epsilon(f(x) - T)\left|f'(x)\right| \mathrm{d}x \\ &= \frac{1}{2\epsilon} \int_c^d \left|f'(x)\right| \mathrm{d}x \end{aligned}$$

Without double roots:

$$\begin{aligned} &= \frac{1}{2\epsilon}\left|\int_c^d f'(x)\,\mathrm{d}x\right| \\ &= \frac{1}{2\epsilon}|f(d) - f(c)| \quad \text{by fundamental theorem of calculus:} \\ &= \frac{1}{2\epsilon} 2\epsilon = 1 \end{aligned}$$

So, we find that summing the integral across the space gives a count for the number of threshold crossings of the process. □

§2.1.1 Expected number of threshold crossing

To find the expected number of threshold crossing events, we want to calculate:

$$\langle N_f\left(\theta\sqrt{r_0}\right)\rangle = \left\langle \int_L \delta\left(f(x) - \theta\sqrt{r_0}\right)\left|f'(x)\right| \mathrm{d}x \right\rangle = \int_L \left\langle \delta\left(f(x) - \theta\sqrt{r_0}\right)\left|f'(x)\right| \right\rangle \mathrm{d}x$$

The expectation in over two variables $(f(x), f'(x))$, which are both Gaussians and jointly distributed. In our case, f has zero mean, and due to constant mean of the process, f' also has zero mean. So, it comes down to calculating the covariance matrix of the two variables to understand their joint distribution. To that end, by virtue of stationarity, note that:

1. $\left\langle f(x)f'(x)\right\rangle = \frac{1}{2}\frac{\partial\left\langle f(x)f(x)\right\rangle}{\partial x} = \frac{1}{2}\frac{\partial}{\partial x} r_0 = 0$

2. $$\begin{aligned} \left\langle f'(x)f'(x)\right\rangle &= \left\langle \left.\frac{\partial f(y)}{\partial y}\right|_{y=x} \left.\frac{\partial f(y)}{\partial y}\right|_{y=x} \right\rangle = \left.\frac{\partial^2\left\langle f(x)f(y)\right\rangle}{\partial x \partial y}\right|_{y=x} \\ &= \left.\frac{\partial^2 r(\Delta x)}{\partial \Delta x \partial \Delta x}\right|_{y=x} \frac{\partial \Delta x}{\partial x}\frac{\partial \Delta x}{\partial y} \\ &= \left.\frac{\partial^2 r(\Delta x)}{\partial \Delta x^2}\right|_{\Delta x=0} \frac{\partial\left(y-x\right)}{\partial x}\frac{\partial\left(y-x\right)}{\partial y} \\ &= -r''(0) = -r_0'' \end{aligned}$$

Together, we can write the joint distribution of $(f(x), f'(x))$ as:

$$\begin{pmatrix} f(x) \\ f'(x) \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} r_0 & 0 \\ 0 & -r_0'' \end{pmatrix}\right)$$

This leads to two observations: first, the expectation is independent of the exact value of x and second, the the process and it's derivative are independent at any given location. Both of these are blessings of stationarity. With this, we can first write:

$$\begin{aligned}\langle N_f\left(\theta\sqrt{r_0}\right)\rangle &= \left\langle \delta\left(f(x)-\theta\sqrt{r_0}\right)\left|f'(x)\right|\right\rangle \int_L \mathrm{d}x \\ &= L\cdot\mathbb{P}\left[f(x)=\theta\sqrt{r_0}\right]\cdot\left\langle\left|f'(x)\right|\right\rangle \\ &= \frac{L}{\sqrt{2\pi r_0}}\exp\left[-\frac{\theta^2}{2}\right]\cdot\sqrt{-\frac{2r_0''}{\pi}} \\ &= \frac{L}{\pi}\sqrt{-\frac{r_0''}{r_0}}\exp\left[-\frac{\theta^2}{2}\right]\end{aligned}$$

The density of excursion sets then is one-half the number of threshold crossings divided by the volume of the space:

$$\mu\left(\theta\sqrt{r_0}\right)=\frac{1}{2}\frac{\left\langle N_f\left(\theta\sqrt{r_0}\right)\right\rangle}{L}=\frac{1}{2\pi}\sqrt{-\frac{r_0''}{r_0}}\exp\left[-\frac{\theta^2}{2}\right] \tag{7}$$

With this we can immediately approximate the mean size of the excursion sets and mean gaps between them. To that end, note that the process spend $L\left(1-\phi(\theta\sqrt{r_0})\right)$, so the mean size s and mean gap $\bar{s}$ can be found as:

$$s=\frac{L\left(1-\phi(\theta\sqrt{r_0})\right)}{L\cdot\mu}=2\pi\left(1-\phi(\theta\sqrt{r_0})\exp\left[\frac{\theta^2}{2}\right]\sqrt{-\frac{r_0}{r_0''}}\right. \tag{8}$$

$$\bar{s}=2\pi\phi(\theta\sqrt{r_0})\exp\left[\frac{\theta^2}{2}\right]\sqrt{-\frac{r_0}{r_0''}} \tag{9}$$

Note that for high threshold, the gap between the excursion is much larger than the size of the excursion set, and in particular, the correlation between such consecutive sets will have decayed close to zero resulting in a Poisson character of the crossings, whose mean is given above, thus full characterizing the distribution.

§2.2 Bochner theorem and spectral density

Theorem 2.2 (Bochner's theorem)

A continuous function $r(x)$, real or complex, is non-negative definite, and hence a covariance function, if and only if there exists a non-decreasing, right continuous, and bounded real function $\rho(\omega)$, such that

$$r(\Delta x) = \int_{-\infty}^{\infty} e^{i\omega\Delta x}\, \mathrm{d}\rho(\omega) \tag{10}$$

In particular, $\rho(\omega)$ has all the properties of measure (after normalization) and is called the spectral density of the process. Then, the k^{th} derivative of a stationary process $\{f(x), x \in \mathbb{R}\}$ when exists has the covariance function

$$r_{f^{(k)}}(\Delta x) = (-1)^k r_f^{(2k)}(\Delta x) = \int_{-\infty}^{\infty} \omega^{2k} e^{i\omega\Delta x}\, \mathrm{d}\rho(\omega) \tag{11}$$

If spectrum is continuous, the spectral density is

$$\rho_{f^{(k)}}(\omega) = \omega^{2k}\rho_f(\omega). \tag{12}$$

In particular it implies:

$$r_0 = \int_{-\infty}^{\infty} \mathrm{d}\rho(\omega) = \omega_0 \quad ; \quad -r_0'' = \int_{-\infty}^{\infty} \omega^2\, \mathrm{d}\rho(\omega) = \omega_2$$

Meaning, we can rewrite the Kac-Rice formula as:

$$\mu\left(\theta\sqrt{r_0}\right) = \frac{1}{2\pi}\exp\left[-\frac{\theta^2}{2}\right]\sqrt{-\frac{r_0''}{r_0}} = \frac{1}{2\pi}\exp\left[-\frac{\theta^2}{2}\right]\sqrt{\frac{\int_{-\infty}^{\infty}\omega^2\,\mathrm{d}\rho(\omega)}{\int_{-\infty}^{\infty}\mathrm{d}\rho(\omega)}} = \frac{1}{2\pi}\exp\left[-\frac{\theta^2}{2}\right]\sqrt{\frac{\omega_2}{\omega_0}} \tag{13}$$

The mean-crossing rate can be found as

$$\mu(0) = \frac{1}{2\pi}\sqrt{\frac{\omega_2}{\omega_0}}, \tag{14}$$

which has units that are inverse of the space and can be thought of as the inverse of the correlation length of the process.

If we repeat the same calculation for the derivative process we will find that the density of maxima and minima of the process as:

$$\mu_+ = \mu_- = \frac{1}{2\pi}\sqrt{\frac{\omega_4}{\omega_2}}; \quad \text{where, } \omega_4 = \int_{-\infty}^{\infty}\omega^4\,\mathrm{d}\rho(\omega) \tag{15}$$

which, when combined with mean-crossing gives the average number of local maxima per mean level crossing,

$$1/\alpha = \frac{\frac{1}{2\pi}\sqrt{\omega_4/\omega_2}}{\frac{1}{2\pi}\sqrt{\omega_2/\omega_0}} = \sqrt{\frac{\omega_0\omega_4}{\omega_2^2}}. \tag{16}$$

The parameter α is invariant under time and scale changes. It is bounded by $0 < \alpha < 1$, and it can be used as an irregularity measure: an α near 1 indicates a very regular process with approximately one local maximum and minimum between mean level upcrossings. If α is near zero the process contains many high frequency component.

§2.3 Structure of the excursion sets

§2.3.1 Threshold derivative

In the study of phenomenology of the Gaussian process, we'd ideally like to go beyond counting excursion sets and try to understand something about the process during the excursion. One approach is to look closely at the process at and near the threshold crossing and generalize the Kac-Rice method to try and gather information about the process in the vicinity of the threshold crossing. More precisely, extension of Kac-Rice based ideas have resulted in in what's called Palm conditioning techniques, simplest of which involves the distribution of the derivatives. The idea is as follows: from the Kac-Rice approach we have already counted all the cases in which the the process crosses a threshold irrespective of the derivative. Now, we will focus on upward crossings, and try to count the upward crossing in which the derivative was restricted as $\#\{f(x) = \theta\sqrt{r_0} \text{ and } 0 < f'(x) < v\}$. Then, we can find cumulative density function of the derivative at the threshold crossing [5].

Theorem 2.3

Derivatives of a stationary Gaussian process at it's up-corssings are distributed as Rayleigh random variable

$$\mathbb{P}_{\theta\sqrt{r_0}}\left(f'\right) = \frac{f'}{\omega_2}\exp\left(-\frac{f'^2}{2\omega_2}\right), \tag{17}$$

and is independent of the threshold.

Proof.

$$\begin{aligned}
\mathbb{P}\left[f(x) = \theta\sqrt{r_0}\,,\, f'(x) < v\right] &= \frac{\#\{f(x) = \theta\sqrt{r_0} \text{ and } 0 < f'(x) < v\}}{\#\{f(x) = \theta\sqrt{r_0} \text{ and } 0 < f'(x)\}} \\
&= \frac{\int_L \left\langle \delta(f(x) = \theta\sqrt{r_0}) f'(x) \mid 0 < f'(x) < v \right\rangle \mathrm{d}x}{\int_L \left\langle \delta(f(x) = \theta\sqrt{r_0}) f'(x) \mid 0 < f'(x) \right\rangle \mathrm{d}x} \\
&= \frac{L \cdot \mathbb{P}\left[f(x) = \theta\sqrt{r_0}\right] \int_0^v f' \exp\left(-\frac{f'^2}{2\omega_2}\right) \mathrm{d}f'}{L \cdot \mathbb{P}\left[f(x) = \theta\sqrt{r_0}\right] \int_0^\infty f' \exp\left(-\frac{f'^2}{2\omega_2}\right) \mathrm{d}f'} \\
&= \frac{\int_0^v f' \exp\left(-\frac{f'^2}{2\omega_2}\right) \mathrm{d}f'}{\int_0^\infty f' \exp\left(-\frac{f'^2}{2\omega_2}\right) \mathrm{d}f'} = \frac{\int_0^v f' \exp\left(-\frac{f'^2}{2\omega_2}\right) \mathrm{d}f'}{\sqrt{2\pi\omega_2} \times \sqrt{\frac{\omega_2}{2\pi}}} \\
&= \int_0^v \frac{f'}{\omega_2} \exp\left(-\frac{f'^2}{2\omega_2}\right) \mathrm{d}f'
\end{aligned}$$

□

But before moving on, let's recall that f and f' are both Gaussian random variable, and furthermore are independent of each other in our case. Then perhaps one would assume that there is nothing to be done since f' will simply be a Gaussian centered at 0. But this is not quite right. This **would** be true if we were measuring the slope distribution deterministically, but instead we are only measuring it in location determined by the process itself, which biases the sampling and results in a different distribution.

§2.3.2 Slepian process

With that, we have all the ingredient we need in order to look more closely into the excursion sets. The idea was firs proposed by Slepian who constructed a model for Gaussian process after crossing a threshold, i.e, a model for the individual excursion after the threshold crossing. It makes use of the fact that at the upward crossing the derivative of the process is distributed as a Ryeligh distributed.

It is also somewhat annoying to constantly write $\theta\sqrt{r_0}$, so without loss of generality, we will simply write θ hereon.

Definition 2.4 (Slepian model)

Let $\{f(x), x \in \mathbb{R}\}$ be a stationary stochastic process and fix a level u, such that $f(x)$ has a finite number of upcrossings in any finite interval. A Slepian model process for $\{f(x), x \in \mathbb{R}\}$ after any $\theta-$upcrossings is any stochastic process $\{\xi_\theta(x), x \in \mathbb{R}\}$ whose finite-dimensional distributions are given by

$$\mathbb{P}\left(\xi_u(x) \leq \mathbf{v}\right) = \int_0^\infty \mathbb{P}_\theta(z)\mathbb{P}\left(f(x) \leq \mathbf{v} \mid f(0) = \theta, f'(0) = z\right) \mathrm{d}z\,, \tag{18}$$

where $\mathbb{P}_u$ is the Rayleigh density.

In the case of a Gaussian process, the conditional distribution of $f(x)$ given $f(0) = u$ and $f'(0) = z$ is still a Gaussian variable, so we simply need the first two moment function to completely characterize the excursion process: $\langle f(x) \mid f(0) = u, f'(0) = z\rangle$ and $\langle f(x)f(y) \mid f(0) = u, f'(0) = z\rangle$.

We take $\boldsymbol{\xi} = (f(y), f(x))\,, \boldsymbol{\eta} = (f(0), f'(0))$, and calculate the joint covariance matrix of $(\boldsymbol{\xi}, \boldsymbol{\eta})$ from the covariance function $r(x)$ for $\{f(x), x \in \mathbb{R}\}$. Note that all the relevant means are 0.

We have four variables, and we have previously identified the covariance between all the pairs except one: $\langle f(x)f'(y)\rangle$. To that end, using Bochner's theorem, we get:

$$\langle f(x)f(y)\rangle = \int \exp\left[i\omega\left(y - x\right)\right] \mathrm{d}\rho(\omega)\,; \quad y > x$$

Differentiating w.r.t x :

$$\begin{aligned}\left\langle f'(x)f(y)\right\rangle &= \int -i\omega \exp\left[i\omega\left(y - x\right)\right] \mathrm{d}\rho(\omega) \\ &= -r'(\Delta x)\end{aligned}$$

Resulting in the following covariance matrix:

$$\boldsymbol{\Sigma} = \left(\begin{array}{cc|cc} r(0) & r(\Delta x) & r(y) & -r'(y) \\ r(-\Delta x) & r(0) & r(x) & -r'(x) \\ \hline r(y) & r(x) & r(0) & 0 \\ -r'(y) & -r'(x) & 0 & -r''(0) \end{array}\right) = \left(\begin{array}{cc} \boldsymbol{\Sigma}_{\xi\xi} & \boldsymbol{\Sigma}_{\xi\eta} \\ \boldsymbol{\Sigma}_{\eta\xi} & \boldsymbol{\Sigma}_{\eta\eta} \end{array}\right).$$

Together we have:

$$\begin{pmatrix}\boldsymbol{\xi}\\ \boldsymbol{\eta}\end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix}\mathbf{0}\\ \mathbf{0}\end{pmatrix}, \left(\begin{array}{cc|cc} r(0) & r(\Delta x) & r(y) & -r'(y)\\ r(-\Delta x) & r(0) & r(x) & -r'(x)\\ \hline r(y) & r(x) & r(0) & 0\\ -r'(y) & -r'(x) & 0 & -r''(0)\end{array}\right)\right) \tag{19}$$

With $\omega_0 = r_0, \omega_2 = -r_0''$ we get the conditional expectation and covariance matrix given $\boldsymbol{\eta} = (\theta, z)$ as

$$\langle \boldsymbol{\xi} \mid \boldsymbol{\eta} = \mathbf{y}\rangle = \boldsymbol{\Sigma}_{\boldsymbol{\xi}\boldsymbol{\eta}}\boldsymbol{\Sigma}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1}\mathbf{y}' = \begin{pmatrix}\theta r(y)/\omega_0 - z r'(y)/\omega_2\\ \theta r(x)/\omega_0 - z r'(x)/\omega_2\end{pmatrix},$$
$$\boldsymbol{\Sigma}_{\boldsymbol{\xi}\xi|\boldsymbol{\eta}} = \boldsymbol{\Sigma}_{\boldsymbol{\xi}\boldsymbol{\xi}} - \boldsymbol{\Sigma}_{\boldsymbol{\xi}\boldsymbol{\eta}}\boldsymbol{\Sigma}_{\eta}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{\eta}\boldsymbol{\xi}} = \begin{pmatrix} r_\kappa(y,y) & r_\kappa(y,x)\\ r_\kappa(x,y) & r_\kappa(x,x)\end{pmatrix}.$$

where,

$$r_\kappa(y,x) = r(\Delta x) - \frac{r(y)r(x)}{\omega_0} - \frac{r'(y)r'(x)}{\omega_2}.$$

This is a covariance structure and it can be used as the covariance function for a non-stationary Gaussian process, where the last two terms are due to conditional induced reduction in covariance. When x or y tend to infinity, the reduction terms go to 0. Similarly $\theta r(x)/\omega_0 - z r'(x)/\omega_2$ is the (random) mean value function. Together, this results in the following theorem:

Theorem 2.5

The Slepian model for a Gaussian process $\{f(x), x \in \mathbb{R}\}$ after upcrossings has the form

$$\xi_\theta(x) = \frac{\theta r(x)}{\omega_0} - \frac{\zeta r'(x)}{\omega_2} + \kappa(x), \tag{20}$$

where ζ has the Rayleigh density $p^u(z) = (z/\omega_2)\, e^{-z^2/2\omega_2}, z \geq 0$, and the residual process $\{\kappa(x), x \in \mathbb{R}\}$ is a non-stationary Gaussian process, independent of ζ, with mean zero and covariance function $r_\kappa\,(x, y)$.

§2.3.3 High excursion

To further study the structure of the excursion set, we will limit our-self to high threshold regime, where we can hope to progress by expanding the Slepian model $\xi_\theta(x)$ in a Taylor series as $\theta \to \infty$. It will turn out that the length and height of the excursion will both be of the order θ^{-1}, so we normalize the scales of $\xi_\theta(x)$ by that factor. First, we have simply

$$r(x/\theta) = \omega_0 - \omega_2\frac{x^2}{2\theta^2}(1 + o(1)), \quad r'(x/\theta) = -\omega_2\frac{x}{\theta}(1 + o(1))$$

as $x/\theta \to 0$. Further, it is easy to see from the covariance function $r_\kappa\,(x, y)$ that $\kappa(x/\theta) = o(x/\theta)$, and we get, omitting all o-terms, and with ζ as the Rayleigh slope variable,

$$\theta\left\{\xi_\theta(x/\theta)-\theta\right\} = \theta\left\{\theta\left(\frac{r(x/\theta)}{\omega_0}-1\right)-\zeta\frac{r'(x/\theta)}{\omega_2}+\kappa(x/\theta)\right\} \tag{21}$$

$$\approx \zeta x - \frac{\omega_2 x^2}{2\omega_0} = -\frac{\omega_2}{2\omega_0}\left(x-\frac{\zeta\omega_0}{\omega_2}\right)^2+\frac{\zeta^2\omega_0}{2\omega_2} \tag{22}$$

Thus, the excursion above a high level θ takes the approximate form of a parabola with height $\frac{\zeta^2\omega_0}{2\theta\omega_2}$ and length $\frac{2\omega_0\zeta}{\theta\omega_2}$. So, the excursion size is distributed as a Rayleigh random variable and the excursion height is distributed as square of Rayleigh distribution. In dimension D higher than 1, in case of isotropic process, the D major lengths of the excursion sets will all be approximately distributed as Rayleigh to the power D.

Denote S as the size of the excursion set. Then S is a Rayleigh random variable with PDF $\mathbb{P}_S(s) = \frac{s}{\omega_2}\exp\left(-\frac{s^2}{2\omega_2}\right)$. We want to find the distribution of $V = S^D$. Using the Jacobian correction, we compute the PDF of V as follows:

$$\begin{aligned}
\mathbb{P}(v) &= \mathbb{P}_S(s)\left|\frac{dS}{dV}\right| \\
&= \mathbb{P}_S(v^{1/D})\cdot\frac{1}{D}v^{\frac{1}{D}-1} \\
&= \frac{v^{1/D}}{\omega_2}\exp\left(-\frac{v^{2/D}}{2\omega_2}\right)\cdot\frac{1}{D}v^{\frac{1}{D}-1} \\
&= \frac{v^{\frac{2}{D}-1}}{D\omega_2}\exp\left(-\frac{v^{2/D}}{2\omega_2}\right).
\end{aligned}$$

In particular, when $D = 2$, we get both the size height distribution of the excursion set as an exponential distribution.

§3 Coding properties

So far, we have looked at the structural and phenomenological properties of Gaussian process in the context of neural code. But now, we will shift to a more functional point of view. Neurons create a representation of the external world which can be read out by other brain areas to query about the information. For example, CA1 cells in hippocampus are widely read out by neurons spread through the brain, and at their simplest, they code for the spatial location of the environment. We have shown that such a code is statistically a threshold crossing of Gaussian process over the space. But what might be the coding properties and benefits of such a code? To answer this we will follow ref [6] in a simplified setting without the threshold and formulate a coding problems in terms of maximum likelihood readout in presence of noise.

Consider a population of N CA1 neurons coding for $1D$ space of length L. Each neurons generates it's code based on independent samples from a stationary Gaussian process $\mathbf{f} = (f_1(x), \ldots, f_N(x))$ with covariance function $r(\Delta x)$. Suppose that the neural representation is further corrupted by an additive Gaussian noise $z_i \sim \mathcal{N}(0, \eta^2)$, so that the activity readout at say $x = 0$ of neuron i is $f_i(0) + z_i$. Then, we formulate the readout problem as the maximum likelihood estimation of the location x given by $\hat{x}$, which in the case of Gaussian noise minimized the mean square error:

$$\hat{x} = \arg\min_x \sum_{i=1}^{N} [f_i(x) - (f_i(0) + z_i)]^2 \tag{23}$$

With this, we'd like to characterize the codding efficiency of the code using a loss function, which we simply take to be the mean squared loss function (we can be broad about it and select other kinds of loss functions without too much overhead!). That is, we'd like to calculate:

$$\mathcal{L} = \left\langle (\hat{x} - x)^2 \right\rangle \tag{24}$$

To that end, we note that we have two separate causes for error in our coding scheme. First, is a local error that occurs due to the resolution imposed by the noise in the code, this occurs when the Bayesian posterior of the estimation has one peak around the true value x. Second, there cal be a global error, where due to the noise, the code at one point the space ends up being "closer" to an entirely different point in the space, and in such cases, we expect more than one peak in the posterior. These two source of error will have to be dealt with separately, and we will start with the one that is more common in computational neuroscience, is easier to handle, and can be done exactly.

§3.1 Local error

Suppose that the error is small Δx such that $\hat{x} = \Delta x$. Then,

$$\begin{aligned}
\hat{x} &= \arg\min_{\Delta x} \sum_{i=1}^{N} [f_i(x) - (f_i(0) + z_i)]^2 \\
&= \arg\min_{\Delta x} \sum_{i=1}^{N} \left[f_i(0) + f_i'(0)\Delta x - (f_i(0) + z_i)\right]^2 \\
&= \arg\min_{\Delta x} \sum_{i=1}^{N} \left[f_i'(0)\Delta x - z_i\right]^2
\end{aligned}$$

Differentiating w.r.t Δx, we get:

$$\sum_{i=1}^{N} f_i'(0)\left[f_i'(0)\Delta x - z_i\right] = 0 \implies \Delta x = \frac{\sum_{i=1}^{N} f_i'(0) z_i}{\sum_{i=1}^{N} f_i'(0)^2}$$

So, we can get expected local error by first taking expectation over the noise:

$$\langle \mathcal{L}_{\text{Local}} \rangle_z = \left\langle (\Delta x)^2 \right\rangle_z = \frac{\eta^2 \sum_{i=1}^{N} f_i'(0)^2}{\left(\sum_{i=1}^{N} f_i'(0)^2\right)^2} = \frac{\eta^2}{\omega_2 \sum_{i=1}^{N} \frac{f_i'(0)^2}{\omega_2}}$$

To avoid long integral, note that $\sum_{i=1}^{N} \frac{f_i'(0)^2}{\omega_2}$ is a Chi-square distribution with degree of freedom N, and the mean of the inverse chi-square distribution of degree of freedom N is $N-2$, resulting in:

$$\langle \mathcal{L}_{\text{Local}} \rangle = \frac{\eta^2}{\omega_2 (N-2)} \tag{25}$$

§3.2 Global error

We will limit ourselves to the case where Global error are rare and thus the place where they can occur are far apart. This will take us to a regime where the numbers of global error can is distributed as a Poisson distribution. Let λ be the rate at which the global errors occur. Then, the probability that there will be any global error is:

$$\mathbb{P}(\text{Global error}) = 1 - \exp(-\lambda L) \approx \lambda L$$

So, we need to estimate λ in order to get a handle on the global error. Global errors occur when the distance between the true location and the noisy code $\|\mathbf{f}(0) - \mathbf{f}(0) + \mathbf{z}\| = \|\mathbf{z}\|$ is larger than the distance between the code at some arbitrary point x (further away than it's correlation length) and the noisy code $\|\mathbf{f}(x) - \mathbf{f}(0) + \mathbf{z}\|$, i.e

$$\|\mathbf{f}(x) - \mathbf{f}(0) + \mathbf{z}\|^2 < \|\mathbf{z}\|^2$$

Define $\widetilde{f} = \|\mathbf{f}(x) - \mathbf{f}(0)\|$ and observe that the inequality boils down to:

$$\mathcal{I} = \left\|\widetilde{f}\right\|^2 - 2\widetilde{f} \cdot z < 0$$

Here, we run into an issue, which is that, if there is a point at which a global error occurs, the error can occur in points that are close to it as well and we run risk of inflating the global error. So, for our purposes here, we will make a simplification which can be understood in either of the two ways: taking the correlation length of the process to be close to 0, or segmenting the process in sizes proportional to the correlation length, such that the segments are essentially independent.

We will proceed in two steps: first averaging over the noise, followed by averaging over the process. Keeping the value of $\widetilde{f}$ fixed, we note that $\mathcal{I}$ is conditionally distributed as $\mathcal{N}\left(\left\|\widetilde{f}\right\|^2, 4\eta^2\left\|\widetilde{f}\right\|^2\right)$. Now, notice that we are interested in the situation when $\mathcal{I} < 0$,

so dividing by a positive constant, such as $\left\|\widetilde{f}\right\|$ will not change the probability of such an event and instead give us an updated $\hat{\mathcal{I}} = \frac{\mathcal{I}}{\left\|\widetilde{f}\right\|}$ which is distributed as $\mathcal{N}\left(\left\|\widetilde{f}\right\|, 4\eta^2\right)$. Then, we can write the probability of global error as:

$$\begin{aligned}
\mathbb{P}\left(\mathcal{I} < 0\right) &= \int_{-\infty}^{0} \frac{1}{\sqrt{8\pi\eta^2}} \exp\left(-\frac{\left(\hat{\mathcal{I}} - \left\|\widetilde{f}\right\|\right)^2}{8\eta^2}\right) \mathrm{d}\hat{\mathcal{I}} \\
&= \int_{-\infty}^{0} \frac{\sqrt{2r_0}}{\sqrt{8\pi\eta^2}} \exp\left(-2r_0 \frac{\left(\hat{\mathcal{I}} - \frac{\left\|\widetilde{f}\right\|}{\sqrt{2r_0}}\right)^2}{8\eta^2}\right) \mathrm{d}\hat{\mathcal{I}} \\
&= \int_{-\infty}^{0} \sqrt{\frac{r_0}{4\pi\eta^2}} \exp\left(-\frac{r_0\left(\hat{\mathcal{I}} - \frac{\left\|\widetilde{f}\right\|}{\sqrt{2r_0}}\right)^2}{4\eta^2}\right) \mathrm{d}\hat{\mathcal{I}}
\end{aligned}$$

With this, we are left with the final task of averaging over the randomness from the process which only appears through $\frac{\left\|\widetilde{f}\right\|}{\sqrt{2r_0}}$ in our case, and the term $\left(\hat{\mathcal{I}} - \frac{\left\|\widetilde{f}\right\|}{\sqrt{2r_0}}\right)^2$ is distributed as a non-central chi square distribution with degrees of freedom N and non-centrality $\hat{\mathcal{I}}$. So, we can take the expectation of the exponential terms in terms of the non-central chi square distribution (again, to avoid an integral), noting that it has the form of a Moment generating function with $t = -\frac{r_0}{4\eta^2}$:

$$\begin{aligned}
&\left\langle \exp\left(-\frac{r_0\left(\hat{\mathcal{I}} - \frac{\left\|\widetilde{f}\right\|}{\sqrt{2r_0}}\right)^2}{4\eta^2}\right)\right\rangle \\
&= \frac{\exp\left(\frac{\hat{\mathcal{I}}\cdot\frac{-r_0}{4\eta^2}}{1+\frac{r_0}{2\eta^2}}\right)}{\left(1+\frac{r_0}{2\eta^2}\right)^{\frac{N}{2}}} = \exp\left(-\frac{\hat{\mathcal{I}}}{2\left(1+\frac{2\eta^2}{r_0}\right)}\right)\left(1+\frac{r_0}{2\eta^2}\right)^{-\frac{N}{2}}
\end{aligned}$$

Putting it together, we get:

$$\begin{aligned}
\mathbb{P}\left(\mathcal{I} < 0\right) &= \sqrt{\frac{r_0}{4\pi\eta^2}} \cdot \sqrt{2\pi\left(1+\frac{2\eta^2}{r_0}\right)} \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi\left(1+\frac{2\eta^2}{r_0}\right)}} \exp\left(-\frac{\hat{\mathcal{I}}}{2\left(1+\frac{2\eta^2}{r_0}\right)}\right)\left(1+\frac{r_0}{2\eta^2}\right)^{-\frac{N}{2}} \mathrm{d}\hat{\mathcal{I}} \\
&= \sqrt{\frac{r_0}{2\eta^2}\left(1+\frac{2\eta^2}{r_0}\right)}\left(1+\frac{r_0}{2\eta^2}\right)^{-\frac{N}{2}} \cdot \frac{1}{2} \\
&= \frac{1}{2}\sqrt{1+\frac{r_0}{2\eta^2}}\left(1+\frac{r_0}{2\eta^2}\right)^{-\frac{N}{2}} \\
&= \frac{1}{2}\left(1+\frac{r_0}{2\eta^2}\right)^{-\frac{N-1}{2}}
\end{aligned}$$

So, all in all, we get the probability of global error, noting that we have effectively space of size $L\omega_2$ since we divide the space into independent parts:

$$\mathbb{P}(\text{Global error}) = \lambda L\omega_2 = \frac{L\omega_2}{2}\left(1+\frac{r_0}{2\eta^2}\right)^{-\frac{N-1}{2}}$$

Finally, the size of the global error is l^2, where l is a uniform random variable over the space $[0, L]$. So, the expected magnitude of the error is $\frac{1}{12}L^2$. So, we get:

$$\langle \mathcal{L}_{\text{Global}} \rangle = \frac{1}{12}L^2\mathbb{P}(\text{Global error}) = \frac{L^3\omega_2}{24}\left(1+\frac{r_0}{2\eta^2}\right)^{-\frac{N-1}{2}}$$

Adn the total expected error combined with the local error is:

$$\langle \mathcal{L} \rangle = \langle \mathcal{L}_{\text{Local}} \rangle + \langle \mathcal{L}_{\text{Global}} \rangle = \frac{\eta^2}{\omega_2(N-2)} + \frac{L^3\omega_2}{24}\left(1+\frac{r_0}{2\eta^2}\right)^{-\frac{N-1}{2}}$$

Note the different role ω_2 plays here (and we can calculate the optimal ω_2 by minimizing the total error)[6]. Also note the exponential suppression of error in the number of neurons.

References

[1] Nischal Mainali, Rava Azeredo da Silveira, and Yoram Burak. *Universal Statistics of Hippocampal Place Fields across Species and Dimensionalities*. June 2024. DOI: 10.1101/2024.06.11.597569. (Visited on 11/15/2024).

[2] A. W. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge: Cambridge University Press, 1998. ISBN: 978-0-521-78450-4. DOI: 10.1017/CBO9780511802256. (Visited on 04/17/2024).

[3] E. L. Lehmann and George Casella. *Theory of Point Estimation*. 2nd ed. Springer Texts in Statistics. New York: Springer, 1998. ISBN: 978-0-387-98502-2.

[4] S. O. Rice. "Mathematical Analysis of Random Noise". In: *The Bell System Technical Journal* 23.3 (July 1944), pp. 282–332. ISSN: 0005-8580. DOI: 10.1002/j.1538-7305.1944.tb00874.x. (Visited on 03/22/2024).

[5] Mark Kac and David Slepian. "Large Excursions of Gaussian Processes". In: *The Annals of Mathematical Statistics* 30.4 (1959), pp. 1215–1228. ISSN: 0003-4851. JSTOR: 2237465. (Visited on 03/22/2024).

[6] Simone Blanco Malerba et al. *Random Compressed Coding with Neurons*. Jan. 2022. DOI: 10.1101/2022.01.06.475186. (Visited on 03/24/2024).