# Point processes and elements of neural coding

Krešimir Josić

Introduction

Neurons communicate largely via spikes (action potentials) which are stereotyped, rapid transients of their membrane potentials. To simplify the statistical analysis of neuronal data and simulations of neuronal populations, the exact shape of each spike is sometimes disregarded, and only the times of their occurrences are recorded. Therefore, the activity of a neuron is sometimes described by a sequence of times at which the neuron has emitted a spike, $s = \{t_0, t_1, \ldots\}$. An abstract representation of a sequence of spikes, also referred to as a *spike train*, can therefore take the form of a collection of vertical bars[1]:
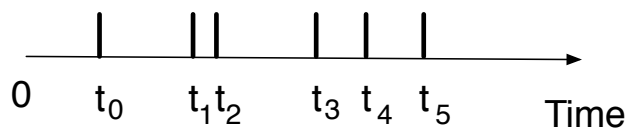
This is an example of what statisticians call a *point process*. Typically, there is some randomness in the spike times. This may be due to the "noisiness" in the response of a cell, or due to other factors that we cannot control or do not know. The goal of the following is to describe the basics of spike train analysis, and give some notion of how neural response can encode information, despite their random appearance.

Basic characterization of a spike train

Modern experimental techniques allow us to record from dozens or more neurons simultaneously, each of which may spike thousands of time. For this data to be meaningful we can look at different statistical descriptions, or try to represent it in a way that tells us about the dynamics and interdependencies between the recorded cells. Let us start with the statistics.

The data consisting of many measurements (shoe sizes of all students at your university, for instance), is usually summarized by its average value. Equally important, however, is the

---

[1]At this point you may ask how to go from the recording of the cell's membrane voltage to a point process. Since spikes do have a width which ranges from somewhere between $.5\ ms$ to $2\ ms$ or more, there is no unique way of doing this. One way is to mark the time at which the membrane voltage exceeds a given threshold as the time of the spike. Since spikes are *rapid* membrane voltage transients, it is also possible to mark the times at which the derivative of the membrane voltage is high.

variance, since it tells you how variable the measurements are. Similar statistics are used to characterize spike trains.

Typical spike trains you may observe in computer simulation or an experiment may look something like this:
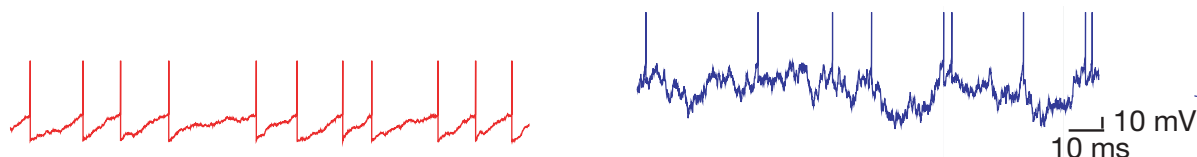
The intensity of the process can be characterized by the *spiking rate*, which we denote r. This can be done by simply counting the number of spikes in a given time window and dividing by the length of the window. The rate tells you how many spikes per unit time are generated, on average, and is typically recorded in $spikes/s = Hz$. Given a constant rate r, the expected number of spikes during a window of length T is simply rT.

Exercise: Approximate the rate of the two processes in the figure above.

The rate, may not be constant over time. This makes the estimation more difficult. However, we start by assuming that the rate is fixed, so that the expected number of spikes in a window of size T is independent of whether this window is at the beginning, middle or end of the spike train. More generally, if the statistics of the process do not change over time, it is called *stationary*.

Another way to get the rate of the process is to start with the intervals between each adjacent pair of spikes in the sequence, $\{T_i\}$. These are called *inter-spike intervals* or *ISIs*. The average ISI, $\bar{T}$, and the rate are inverses of each other (Convince yourself that this is true). Each T_i can be thought of as a realizations of a radom variable T, if the process is stationary.

Note that the two processes in the figure above have approximately equal rates. However, the spikes in the left process appear to be spaced more regularly than on the right. There are a couple of ways to measure the irregularity of the spike train, that is, the variability of the ISIs. The simples measure of variability is the *coefficient of variation* or CV of the ISI distribution. This is defined as the ratio of the standard deviation of the ISIs and their mean[2]:

$$CV = \frac{\sqrt{\frac{1}{N}\sum_i(T_i - \bar{T})^2}}{\bar{T}}.$$

This may not be the best measure of variability, since it does not take into account dependencies between successive ISIs. The *Fano factor* provides an alternative: Let $N(T)$ be the

---

[2]For technical reasons, one actually uses $\sqrt{\frac{1}{N-1}\sum_i(T_i - \bar{T})^2}$ to estimate the standard deviation from a sample of ISIs. This is due to the fact that we think of the measured T_i's as samples from an underlying distribution. To make sure that we have an unbiased estimate of the standard deviation, we use $N - 1$ in the denominator.

The reason we divide by the mean interspike interval is the following: Think of two stocks, say Google, valued at about $450 per share and General Electric at $32 per share. A fluctuation of $1 in the Google stock is much smaller than a fluctuation of $1 in the GE stock. To compare the two we therefore need to normalize.

number of spikes in a window of width T for a stationary process. The Fano Factor is defined as

$$F = \frac{\text{var}[N(T)]}{\langle N(T) \rangle}.$$

Exercise: Show that both F and CV are 1 for the Poisson process.

For theoretical purposes, we think of the ISI as a random variable with probability density $p(t)$. In other words the probability that a spike at t_0 is followed by a spike between $t_0 + T$ and $t_0 + T + \Delta t$ is[3] $p(t)\Delta t + o(\Delta t)$. Note that $p(t)$ is zero for negative values of t.

Exercise: Neurons have a limit to how fast they can fire action potentials. A spike is typically followed by a short period during which another spike is less likely to occur. In an ideal situation we talk of an *absolute refractory period* – an interval of time following a spike during which another spike is impossible.

Compare a neuron with ISI probability density $p(t)$ with a neuron with refractory period t_r by comparing the CV of the distribution $p(t)$ and $p(t - t_r)$. In particular find an expression for CV_r, the coefficient of variation for the second distribution, in terms of CV, the coefficient of variation of the first. What happens as $t_r \to \infty$? Explain this intuitively.

Frequently one is interested in how a neuron, or collection of neurons, respond to a stimulus which may be brief or ongoing, but variable. In such cases (which are, probably, more representative of what is going on in neural tissue), the rate is not constant. We can define the *instantaneous firing rate*, $r(t)$, indirectly by requiring that the probability of a spike occurring between t and $t + \Delta t$ is approximately $r(t)\Delta t$ for small Δt.

The Poisson process and renewal processes

In this section we turn to the question of how to simulate some simple point processes. If the ISIs are independent of each other, the point process is called *renewal*. After the previous section you probably already know how to simulate such a process: Take a distribution $p(t)$ such that $p(t) > 0$ only when $t > 0$. Start with a spike at time $t = 0$, and choose a number t_1 from the distribution $p(t)$. Place, the second spike at t_1, and repeat the process[4].

The above is a perfectly valid way of generating a spike train. The following alternative is computationally more expensive, but offers a completely different view of point processes. Let us assume that we have a process such that

1. The probabilities of a spike occurring during an interval $[t_1, t_1 + s_1]$ and $[t_2, t_2 + s_2]$ are independent, if the intervals are not overlapping.

---

[3]It will be useful to review "little-oh" and "big-oh" notation at this point.

[4]If you don't put a spike at 0, it is a bit more difficult to decide where to put the first spike. This is equivalent to the following problem: suppose you start observing a neuron that has been active for a while. You don't know how long it has been since the last spike. The time at which you observe the first spike is a random variable – What is the corresponding distribution?

2. The probability that a spike occurs during a brief interval of time interval of length Δt is $r\Delta t + o(\Delta t)$.

3. The probability of more than one spike occurring during a time interval of length Δt is $o(\Delta t)$.

Here r is the rate of the process – you can check that the expected number of spikes during an interval T is exactly rT.

Such a process is very easy to simulate approximately. Just split the timeline into bins of some small width Δt, and place a spike into each bin with probability $r\Delta t$. It doesn't even matter in which order you fill the bins, as long as you visit every bin only once.
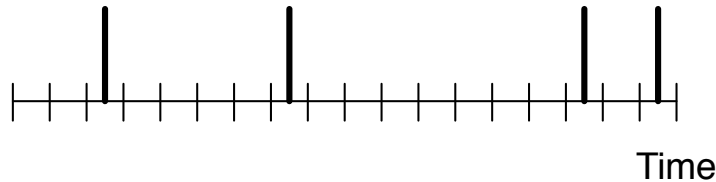
Time

We can derive the ISI probability density for a Poisson process directly from this definition. Note that if a spike occurred at time 0, then the probability that no spike occurred between 0 and t is the probability that no spike occurred in any of the $t/\Delta t$ bins of size Δt subdividing $[0, t]$. This probability is approximately

$$p_0(t) \approx (1 - r\Delta t)^{t/\Delta t}.$$

The approximation becomes better as $\Delta t \to 0$, and in this limit we obtain $p_0(t) = \exp(-rt)$.

The probability that the next spike occurs between t and $t + \Delta t$ is $p(t)\Delta t$, where $p(t)$ is the probability density of the ISI. This is the product of the probability $r\Delta t$ of a spike in this interval, and the probability that no spike occurred up to time t. Therefore

$$p(t)\Delta t = e^{-rt}r\Delta t, \text{ and so } p(t) = re^{-rt}.$$

Exercise: Use this fact to prove that the Poisson process is memoryless. This means that the probability of a spike happening during an interval $[t, t + \Delta t]$ is independent of the time of any previous spike.

Exercise: Show that the probability of k spikes in a time interval T is given by the Poisson distribution. In particular, let N_T be the number of spikes in an interval of length T, then

$$P(N_T = k) = \frac{e^{-rT}(rT)^k}{k!}.$$

While simulating a Poisson process in this way is not very efficient, it has the advantage that it is easily generalized. Sometimes the probability $r(t)\Delta t$ of a spike during a short time interval Δt is known (at least approximately). In this case, the process can be simulated using the 3 steps discussed above. If this probability depends on other spike trains (other neurons in a network), it may be impossible to find a corresponding ISI distribution.
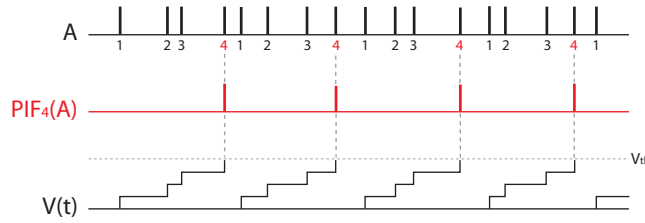
A very simple model of a neuron is provided by the perfect integrate and fire model. Such a neuron is described by a membrane potential $V(t)$. When this membrane potential reaches a threshold voltage V_{th}, the neuron fires an action potential and is reset to some voltage V_r. In the simplest case, the input to such a neuron is a point process itself, and each spike raises the membrane potential $V(t)$ by a fixed amount[5]. If N inputs will cause the neuron to spike, we denote this model PIF_N. In particular, we can think of this as an operation on spike trains. In the figure above you can see the effet of an input (spike train A) on the membrane voltage $V(t)$, and the output spike train $PIF_4(A)$.

> **Exercise:** If the input to a perfect integrate and fire neuron is a Poisson spike train, show that the ISI of the output is a gamma function. Hint: What is the distribution of the sum of i.i.d. random variables that follow the Poisson distribution.

Autocorrelation and the spectral density

The coefficient of variation is a single number, and thus cannot fully capture the statistics even of stationary random processes. It is sometimes more useful to consider a function that gives the probability of a spike at time $t + T$ given that a spike has occurred at time t. As we will see below, this function is closely related to the variance of spike counts N_t.

Assume that we are looking at a stationary process and denote by $N(t, t+T)$ the number of spikes in the interval $(t, t+T)$. Since the process is stationary, we can start our observation at any time, and call that time $t = 0$. We can now make the notion of a conditional spike probability more precise by defining [1]

$$h(t) = \lim_{\Delta t \to 0^+} \frac{Pr[N(t, t + \Delta t) = 1 | N(-\Delta t, 0) > 0]}{\Delta_t}.$$

Note that this is normalized so that $h(t)\Delta t$ can be interpreted as the probability of a spike in the interval $[t, t + \Delta t]$ given that there is a spike at time 0. Also, you can check that $h(t) = h(-t)$. By stationarity, this is the same as the probability of a spike in the interval $[T + t, T + t + \Delta t]$ given a spike at time T.

The conditional intensity $h(t)$ is very useful by itself, since it provides very useful information about several aspects of the spike train. A refractory period is reflected in a small value of $h(t)$ around $t = 0$. Oscillations in the output of a neuron are also easier to see using $h(t)$ (or its Fourier transform) rather than the original spike train.

---

[5]This is a simple model of a neuron with excitatory inputs. Inhibitory inputs can be modeled by a fixed decrease in the membrane potential.

To relate the function $h(t)$ to the variance of a function, first note that

$$\text{var}[N(t, t + \Delta t)] = E[N(t, t + \Delta t)^2] - (E[N[t, t + \Delta t])^2 \tag{1}$$

$$= Pr[N(t, t + \Delta t) = 1] - (Pr[N(t, t + \Delta t) = 1)^2 + o(\Delta t) \tag{2}$$

$$= r\Delta t + o(\Delta t). \tag{3}$$

The equalities follow from the fact that the probability of having a spike in an interval of length Δt is $r\Delta t$, and the probability of having more than one spike is $o(\Delta t)$. Similarly, we find that

$$\text{cov}[N(t, t + \Delta t), N(t + T, t + T + \Delta t)] = rh(T)(\Delta t)^2 - r^2(\Delta t)^2 + o(\Delta t^2) \tag{4}$$

Exercise: Check this relation.

Let us look at a spike count over an interval $(0, t)$ which we divide into $M = t/\Delta t$ pieces of size Δt. Then

$$N(0, t) = \sum_{i=0}^{M} N(i\Delta t, (i + 1)\Delta t).$$

Using the fact that the variance of a sum of random variables equals to the sum of all their variances and covariances, we can see that

$$\text{var}[N(0, t)] = \sum_{i=0}^{M-1} \text{var}[N(i\Delta t, (i+1)\Delta t)] + 2 \sum_{i=1}^{M-1} \sum_{j=1}^{M-i-1} \text{cov}[N(i\Delta t, (i+1)\Delta t), N((i+j)\Delta t, (i+j+1)\Delta t)].$$

The factor of 2 comes from the symmetry of the covariance matrix. Using the facts we derived about the covariance and variance, we see that as Δt approaches zero, this sum approaches[6]

$$\text{var}[N(0, t)] = \int_0^t r \, dz + 2 \int_0^t \int_0^{t-z} (rh(u) - r^2) du \, dz = rt - r^2 t^2 + 2r \int_0^t (t - u)h(u) du. \tag{5}$$

This is a bit complicated, but we can simplify it a bit by introducing the *autocorrelation function*[7], $c(t)$. This function is defined by the following requirement

$$c(t)(\Delta t)^2 + o(\Delta t^2) = \text{cov}[N(t, t + \Delta t), N(t + T, t + T + \Delta t)],$$

and can thus also be thought as the covariance density. To get the variance of $N(t)$, we just split up the interval $(0, t)$ into M pieces again and sum up all the variances. Taking the limit as $\Delta t \to 0$ we again get an integral. Note that the expressions Eq. (1–4) imply

$$c(t) = r\delta(t) - r^2 + rh(u),$$

where $\delta(t)$ is the Dirac delta function. You will also find the autocorrelation of a spike train

---

[6]Remember the definition of a Riemann integral!

[7]This should really be called the auto-covariance function, but the term autocorrelation is encountered more frequently.

$\rho(t)$ expressed as

$$c(t) = \frac{1}{T} \int_0^T E[(\rho(t) - r)(\rho(t + \tau - r)]dt.$$

Exercise: Compute $h(t)$ and the autocorrelation function for a Poisson process.

Multivariate point processes

Our brains consist of many neurons working together. Therefore, we frequently need to extend the methods described above to collections of spike trains. If neurons acted independently, this would be easy since the response of a population of N neurons would be completely characterized by the individual responses of each group in the population.

Neuronal responses are frequently not independent. A group of cells could be coupled, or they could receive shared input. In both cases observing a spike in one cell tells us something about the probability of observing a spike in another. This can again be summarized by using an intensity function. Suppose we have two cells, A and B. Let $N_A(s,t)$ be the number of spikes in cell A during the interval (s,t). The probability of observing a spike in A in the interval $(t, t + \Delta t)$ given a spike in cell B at time 0 is given in terms of the cross-intensity function

$$h_{A,B}(t) = \lim_{\Delta t \to 0^+} \frac{Pr[N_A(t, t + \Delta t) = 1 | N_B(-\Delta t, 0) > 0]}{\Delta t}.$$

It is easy to see that $h_{A,B}(t) = h_{B,A}(-t)$ when the rates of the two processes are equal. It is also easy to see from the definition how to compute $h_{B,A}(t)$: Separate the timeline into bins of some size. Choose a spike in train A at some time t_0 (dashed vertical line in the figure below). Now compute the relative times between this spike and all spikes in train B (in the figure below t_i stands for the relative time between this and the ith spike following it). Increase the value in the bin by this amount.

Spike train A

Spike train B
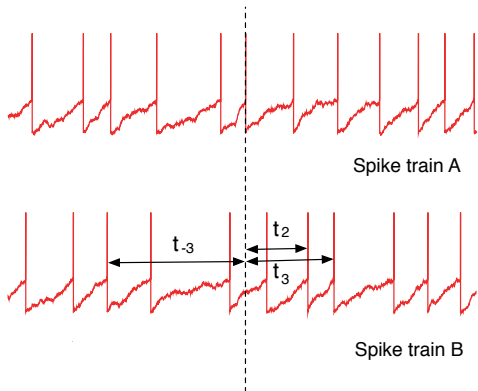
Exercise: How do you need to normalize the resulting histogram to get $h_{B,A}(t)$?

The argument used in proving relation Eq. (4) can again be used to relate the cross-intensity function with the cross-covariance density $c_{A,B}(t) = r_B h_{A,B}(t) - r_A r_B$, where r_A and r_B are the respective rates of the two processes. Here $c_{A,B}(T)(\Delta t)^2$ can again be interpreted as the approximate covariance of $N_A(t + T, t + T + \Delta t)$ and $N_B(t, t + \Delta t)$.

The covariance of the spike counts $\text{cov}(N_A(T), N_B(T))$ is again obtained by an integral of the cross-correlation function $c_{A,B}(t)$

$$\text{cov}(N_A(T), N_B(T)) = \int_{-T}^{T} (T - |t|)c_{AB}(t)dt.$$

The correction factor $T - |t|$ appears because the sliding windows over which the count is performed have an overlap that decreases as they are shifted with respect to each other. A derivation is provided in the Appendix.

The autocorrelation function discussed in the previous section can be identified as $c_{A,A}(t)$, and is therefore a special case of the cross-correlation function.
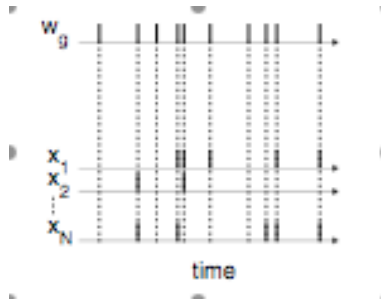

Exercise: Start with a Poisson process and create N daughter processes as follows. For each process i define a probability r_i. Spikes in the process i occur with probability r_i at the same time as those in the mother process (see the Figure above – this figure comes from [3]). Compute the cross-correlation between any two daughter spike trains.

Coding and Fisher information

In the following I will briefly discuss how the concepts introduced so far can be used to examine how information is represented in the collective activity of a population of neurons. I will try to motivate most concepts, however you can find full justifications and omitted details in most books on statistics and signal analysis, such as [2].

Let us look at a visual stimulus consisting of moving bars with a certain orientation. Many neurons in the visual cortex have a preference for certain orientations, and will respond most vigorously to moving bars of this direction. The situation is illustrated schematically in the figure below – a stimulus with orientation 135° is presented on the left generating a response $(20\ Hz)$ that is weak compared to the response to the preferred, horizontal stimulus $(60\ Hz)$.
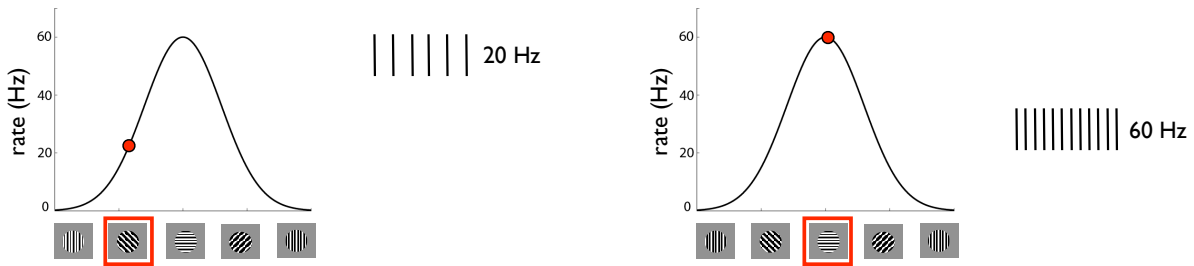
For simplicity we will assume that it is only the rate of the response that carries information about the stimulus[8]. That is, we only need to count the number of spikes during a some time interval of length T to capture all the information about the stimulus available in the response of this neuron.

A simple encoding model is now provided by $r = f(s)$, where r is the rate of the response, s is the stimulus. The graph of f is frequently called a *tuning curve*. The function f could be a gaussian, or anything else that is bell shaped. In this case if one observes the neuron firing at $20Hz$ in the example above, it is immediate that the stimulus s must have been a grating at angle 135° or 45°. To remove this ambiguity, let us assume that, as long as we discuss single cells, only stimuli between 0° and 90° are presented.

Unfortunately, things are not really this simple. Neurons are inherently noisy: If the same stimulus is presented twice, the response will not be exactly the same. If different stimuli are presented, one sees something that looks more like this (Figure from [4] which I recommend if you want to get a better overview of this topic.)
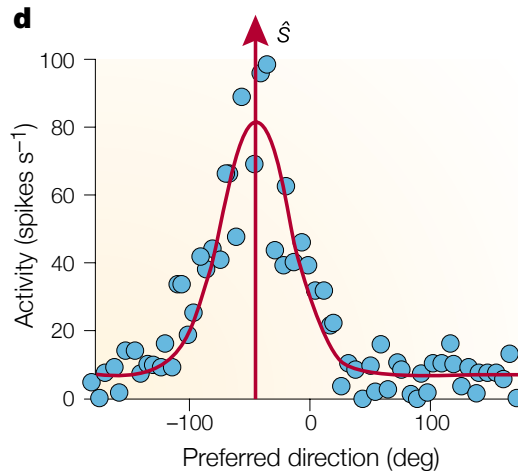
Since the response is variable, it is far better to model the response as

$$r = f(s) + \xi,$$

where ξ is some random variable that may, or may not depend on the stimulus. For instance, ξ could follow a uniform distribution with mean 0 and with variance that increases with the

---

[8]This is called a *rate code*. If information about a stimulus is provided by the spike times of the individual spikes in the train we talk about a *timing code*. It is very likely that, at least in some situations, the timing of individual spikes is important.

intensity of the response.

If we only know the rate of the response, there is a continuum of possibilities for the value of the stimulus. However, some guesses, or estimates, are clearly better than others. In statistics one frequently chooses a systematic way of making a guess, and calls it an *estimator*. If $E[\xi] = 0$ then a good estimator of the stimulus[9], given a response, r could be $s_{est} = f^{-1}(r)$.

Since the estimator is a function of r, it must itself be a random variable that depends on the stimulus. If the expected value of the estimator equals the stimulus, that is $\langle s_{est} \rangle = s$, we say that the estimator is *unbiased*. An estimator that gives the right answer on average may still do a bad job. What we really want is that the estimator s_{est} is consistently close to s. Note that for an unbiased estimator, the variance of the error is just the variance of the estimator

$$\langle (s_{est} - s)^2 \rangle = \langle (s_{est} - \langle s_{est} \rangle)^2 \rangle = \text{var}(s_{est})$$

Therefore, a good unbiased estimator has small variance.

We can say that the response of a cell is as accurate as the best possible unbiased estimate that we can get about the stimulus from observing the output of the cell. Fortunately, this bound can actually be computed. Let $p[r|s]$ be the conditional probability density for the response rate, given a stimulus s. Assuming that $P[r|s]$ is twice differentiable as a function of s, we can define the *Fisher information*

$$I_F(s) = \left\langle \frac{d^2}{ds^2} \log P[r|s] \right\rangle.$$

The inverse of the Fisher information, $1/I_F(s)$, provides a lower bound on the variance of an unbiased decoding estimate of s from the response [2].

We have already established that the nervous system of most animals represents information with more than one cell. In the visual cortex there will be a number of cells responding to the stimulus. Each of these cells may have a different preferred orientation. The collection of tuning curves from a population of cells may look something like this (Figure again from [4])
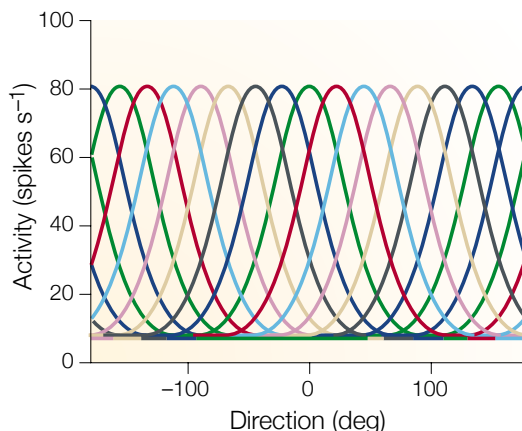
Although this is an idealized model, it does capture one important aspect of the real situation: the average response of each neuron to the stimulus could be different. We can

---

[9]Remember that we are still working under the assumption that only stimuli between 0° and 90° are presented

therefore model the response of a population of neurons to a stimulus s as

$$r_i = f_i(s) + \xi_i, \ i = 1, \dots N, \qquad \text{or, using vector notation,} \qquad \mathbf{r} = \mathbf{f}(s) + \xi.$$

We can again try to estimate the stimulus from the reponse $\mathbf{r}$ of the entire population. The information about the stimulus captured in this response can again be quantified by the minimal variance of an unbiased estimator s_{est}. As above, this bounded below by the inverse of the Fisher information

$$I_F(s) = \left\langle \frac{d^2}{ds^2} \log P[\mathbf{r}|s] \right\rangle.$$

Here $P[\mathbf{r}|s]$ is the conditional probability density for the population response, given a stimulus.

This is only the beginning of the story, and I will talk about several possible further directions during my introductory lecture. In particular, one can ask about the impact of correlations between the responses r_i of the different neurons on the information encoded by the population. We will also discuss how uncertainty about a stimulus can be represented by a population, and why it is better to have many neurons rather than one.

Appendix: Cross-correlation and covariance relation

By a slight abuse of notation, the spike train generated by neuron A will be called A as well. Therefore, A is a sequence of times. This spike train is somtimes represented as a sum of delta functions

$$N'_A(t) = \sum_{t_i \in A} \delta(t - t_i).$$

Note that $N_A(T) = \int_s^t N'_A(\tau)d\tau$, since the integral of each delta function gives a contribution of 1. We therefore have

$$
\begin{aligned}
\text{cov}(N_A(T), N_B(T)) &= \text{cov}\left(\int_0^T N'_A(t)dt, \int_0^T N'_B(s)ds \right) \\
&= \int_0^T \int_{-t}^{T-t} \text{cov}\left(N'_A(t), N'_B(t+\tau) \right) d\tau dt \\
&= \int_{-T}^0 \int_{-\tau}^T \text{cov}\left(N'_A(t), N'_B(t+\tau) \right) dt d\tau \\
&\quad + \int_0^T \int_0^{T-\tau} \text{cov}\left(N'_A(t), N'_B(t+\tau) \right) dt d\tau \\
&= \int_{-T}^T \int_{\max(0,-\tau)}^{\min(T,T-\tau)} \text{cov}\left(N'_A(t), N'_B(t+\tau) \right) dt d\tau \\
&= \int_{-T}^T \int_{\max(0,-\tau)}^{\min(T,T-\tau)} \text{cov}\left(N'_A(0), N'_B(\tau) \right) dt d\tau
\end{aligned}
$$

$$= \int_{-T}^{T} \text{cov}\left(N_A'(0), N_B'(\tau)\right) \left(\int_{\max(0,-\tau)}^{\min(T,T-\tau)} dt\right) d\tau$$

$$= \int_{-T}^{T} \text{cov}\left(N_A'(0), N_B'(\tau)\right) (T - |\tau|) d\tau$$

$$= \int_{-T}^{T} (T - |t|) c_{AB}(t) dt$$

It is frequently more useful to normalize the covariance by the variances.

$$C_T(A, B) = \frac{\text{cov}(N_A(T), N_B(T))}{\text{var}(N_A(T))\text{var}(N_B(T))} \tag{6}$$

By positive definiteness of the covariance matrix, $-1 \leq C_T(A, B) \leq 1$. By the derivation above we have

$$C_T(A, B) = \frac{T \int_{-T}^{T} \left(1 - \frac{|t|}{T}\right) R_{AB}(t) dt}{\sqrt{T^2 \int_{-T}^{T} \left(1 - \frac{|t|}{T}\right) R_{AA}(t) dt \int_{-T}^{T} \left(1 - \frac{|t|}{T}\right) R_{BB}(t) dt}}$$

$$= \frac{\int_{-T}^{T} \left(1 - \frac{|t|}{T}\right) R_{AB}(t) dt}{\sqrt{\int_{-T}^{T} \left(1 - \frac{|t|}{T}\right) R_{AA}(t) dt \int_{-T}^{T} \left(1 - \frac{|t|}{T}\right) R_{BB}(t) dt}}.$$

In the limit as $T \to \infty$ we obtain the correlation coefficient between the two spike trains

$$C_{AB} = \frac{\int_{-\infty}^{\infty} R_{AB}(s) ds}{\sqrt{\int_{-\infty}^{\infty} R_{AA}(s) ds \int_{-\infty}^{\infty} R_{BB}(s) ds}},$$

Thanks go to Robert Rosenbaum who simplified this derivation.

References

[1] DR Cox and V. Isham. *Point Processes*. Chapman & Hall/CRC, 1980.

[2] Steven M. Kay. *Fundamentals of statistical signal processing: estimation theory*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1993.

[3] A. Kuhn, A. Aertsen, and S. Rotter. Higher-Order Statistics of Input Ensembles and the Response of Simple Model Neurons. *Neural Computation*, 15(1):67–101, 2003.

[4] A. Pouget, P. Dayan, and R. Zemel. Information processing with population codes. *Nat Rev Neurosci*, 1(2):125–32, 2000.