Problem Set 6:

Stochastic Theory of Spike Production

Week 9, SoSe 2022 — Benoit Scholtes
Computational Neuroscience — & Stefan Rotter

In the last sheet, we looked at how we could use stochastic processes, namely shot noise, to model the random walk of a neuron's membrane potential in response to the integration of input spikes. This week, we will look at how we can use stochastic theory to model and understand a neuron's production of spikes.

Problem 1: Rolling a dice — 5 points

To motivate what comes next, first observe what happens to the distribution of dice roll averages as you roll more and more dice. In other words, observe how the distribution of $\bar{X}_i = (X_1 + \ldots + X_n)/n$ behaves as $n \to \infty$, for X_i the random variable sampling from the discrete uniform distribution over the space $\{1, 2, 3, 4, 5, 6\}$ (a dice). Plot the distribution (histogram of means obtained) of $\bar{X}_i$ for 2 until 10 dice. What do you notice about how the distribution (and the moments of the distribution) changes as you increase the number of dice? Can you explain how you would calculate this evolution of distributions? You don't actually need to do the calculations.

Please don't waste an evening doing this problem with real dice and excel, use Python. You can however construct the distributions by hand by looking at the number of combinations that give the different averages, without doing experiments.

Solution.

Have a play around with the python code uploaded in ILIAS. Adding *independent* random variables together to form a new random variables $Z = X + Y$, means that Z is distributed as the convolution of the distributions of X and Y: $f_Z(z) = (f_X * f_Y)(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x) \mathrm{d}x$, generalising to any k-number of independent variables: $f_Z(z) = (f_{X_1} * \cdots * f_{X_n})(z)$. Here's the wikipedia entry if you're interested.

For two dice, one can already make sense of this by observing that there are more ways to obtain an average in the middle, around 3 (or sum around 6 and 7), than an average around the ends, of 1 or 6. Namely, obtaining an average of 1 can only be obtained by throwing two 1s, while an average of 3 can be obtained with throws 1 & 5, 2 & 4, 3 & 3, 4 & 2, and 5 & 1, making it 5 times more likely. If one does this for all averages, you obtain a symmetrical triangular distribution centred at 3.5. This distribution is just the convolution of two identical uniform distributions, which is the distribution for the result of rolling a single dice. This should make sense as all the possibilities to obtain an average of $\bar{X}_i$ is given by summing over all such possibilities (here for two dice):

$$\mathbb{P}[\bar{X}_2 = z] = \sum_i \mathbb{P}[X_i = x, Y = 2z - x] = \sum_i \mathbb{P}[X_i = x]\, \mathbb{P}[Y = 2z - x] = (f_X * f_Y)(z),$$

where the second last equality holds because the first roll X and the second roll Y are *independent* random variables. Note that as you add more and more dice, the effect of more combinations of rolls giving averages near the middle is more pronounced, leading to distributions with increasingly smaller tails. This is just given by the k-fold convolution of the uniform distribution for the single dice.

Indeed, you should see in your experiments that rolling any k-number of dice more and more times causes the histogram of means to tend to the distribution obtained from the convolution (shifted to be centred at the mean). Increasing the number of dice k, you should see that the distributions tend roughly towards a normal distribution. Sometimes many rolls are needed to make these distributions representative, unfortunately. You can, however, plot the value of the first few central moments of order great than 2 as more dice are rolled and see that the moments tend to zero (remember the normal distribution can be defined as the only distribution with vanishing central moments greater than 2).

Problem 2: The central limit theorem — 8 points

We can make sense of the last question by appealing to the central limit theorem:

Let $\{X_1, ..., X_n, ...\}$ be independent and identically distributed random variables, each with mean μ and variance σ^2. Due to the law of large numbers, the mean of the sample $\bar{X}_n \equiv (X_1 + ... + X_n)/n$ converges to μ in the $n \to \infty$ limit.
The theorem describes the fluctuations of the sample mean $\bar{X}_n$ about μ during this convergence. It says that in the $n \to \infty$ limit, the distribution of the sample average tends to the normal distribution with mean μ and variance σ^2/n.

Note, this theorem is so useful and significant because it holds irrespective of the distribution of the random variables X_i, as long as it has finite variance. We would like to prove this theorem, given its ubiquity in probability theory. Try to do so by appealing to the characteristic function of the random variables

$$Z_n \equiv \frac{X_1 + ... + X_n - n\mu}{\sqrt{n\sigma^2}} \equiv \sum_{i=1}^{n} \frac{1}{\sqrt{n}} Y_i,$$

where $Y_i = (X_i - \mu)/\sigma$ are newly defined random variables. For reference, the characteristic function of a Gaussian is $\hat{f}(t) = e^{-\mu t} e^{-(\sigma t)^2/2}$ for mean μ and variance σ^2.

Hint: Use the result that having finite k^{th} order moment means that the characteristic function has a k^{th} order derivative, allowing you to expand the expression for the characteristic function using Taylor's theorem. Also, you might find this property of variance useful for relating the new random variable Z_n to the mean $\bar{X}_n$: $\mathbb{V}ar[aX + bY] = a^2\mathbb{V}ar[X] + b^2\mathbb{V}ar[Y] - 2ab\, Cov[X, Y]$, for constants a, b and random variables X, Y.

Solution.

Note that in the limit $n \to \infty$, the sum $X_1 + ... + X_n$ has mean $n\mu$, while the variance, using the property given in the hint, is $n\sigma^2$. Thus, Z_n has mean 0 and to calculate the variance, one can first note that the variance of a random variable plus a constant is just that of the random variable (again from the property in the hint). Thus, the variance of Z_n is the same as that of $\frac{X_1+...+X_n}{\sqrt{n\sigma^2}}$, which from the hint is

$$\mathbb{V}ar\left[\sum_{i=1}^{n}(n\sigma^2)^{-1/2}X_i\right] = n\cdot(n\sigma^2)^{-1}\mathbb{V}ar[X_i] = 1,$$

given that the covariance between random variables X_i is zero due to independence.

This all means that we expect that Z_n tends to a distribution of mean 0 and variance 1 as $n\to\infty$. Now we just want to show that this distribution is the normal distribution $\mathcal{N}(0,1)$, which would be the same as showing that the distribution of the mean $\bar{X}_n$ converges to $\mathcal{N}(\mu,\sigma^2/n)$.

Using the new random variable Y_i, the characteristic function of Z_n is

$$\begin{aligned}\hat{f}_{Z_n}(t) &= \hat{f}_{\sum_{i=1}^{n}Y_i/\sqrt{n}}(t)\\ &= \hat{f}_{Y_1/\sqrt{n}}(t)\cdots\hat{f}_{Y_n/\sqrt{n}}(t)\\ &= \hat{f}_{Y_1}(t/\sqrt{n})\cdots\hat{f}_{Y_n}(t/\sqrt{n})\\ &= \left[\hat{f}_{Y_i}(t/\sqrt{n})\right]^n,\end{aligned}$$

where in the second line we used the identity that the characteristic function of the sum of multiple independent random variables is the product of the individual characteristic functions (conversely, the pdf of the sum is the convolution of the individual pds). In the third line, we used the definition of the characteristic function $\hat{f}_X(t)=\mathbb{E}[e^{itX}]$, and in the last line, that all Y_i are identically distributed and thus have the same characteristic function.

Now we can sub in the definition of the characteristic function and expand the exponential $g(t/\sqrt{n}) = e^{itY_i/\sqrt{n}}$ about $t/\sqrt{n}=0$ using Taylor's theorem:

$$\begin{aligned}\hat{f}_{Z_n}(t) &= \left[\mathbb{E}[e^{itY_i/\sqrt{n}}]\right]^n\\ &= \left[\mathbb{E}\left[g(0)+g'(0)\left(\frac{t}{\sqrt{n}}-0\right)+\frac{g''(0)}{2}\left(\frac{t}{\sqrt{n}}-0\right)^2+o\left(\frac{t^2}{n}\right)\right]\right]^n\\ &= \left[\mathbb{E}\left[1+iY_i\cdot e^{iY_i\cdot 0}\cdot\left(\frac{t}{\sqrt{n}}\right)+\frac{i^2Y_i^2\cdot e^{iY_i\cdot 0}}{2}\cdot\left(\frac{t^2}{n}\right)+o\left(\frac{t^2}{n}\right)\right]\right]^n\\ &= \left[1+i\frac{\mathbb{E}[Y_i]}{\sqrt{n}}\cdot t-\frac{\mathbb{E}[Y_i^2]}{2n}\cdot t^2+o\left(\frac{t^2}{n}\right)\right]^n\\ &= \left[1-\frac{t^2}{2n}+o\left(\frac{t^2}{n}\right)\right]^n,\end{aligned}$$

where $o(\frac{t^2}{n})$ is little-o notation which indicates an error term (or just function) that decays *faster* than t^2/n in the limit $t^2/n\to\infty$. This is just from Taylor's theorem, as all subsequent terms in the series are functions of higher orders of $t/\sqrt{n}$. In the last line, we used that $\mathbb{E}[Y_i]=0$ and $\mathbb{E}[Y_i^2]=\mathbb{V}ar[Y_i]+\mathbb{E}[Y_i]^2=1$.

We can now use the limit $n \to \infty$ and note that the higher order terms $o(\frac{t^2}{n})$ vanish in the limit and that the limit of the exponential function is $e^x = \lim_{n\to\infty}(1 + x/n)^n$, such that

$$\hat{f}_{Z_n}(t) \to e^{-t^2/2}, \quad n \to \infty$$

This is the characteristic function of a Gaussian for mean $\mu = 0$ and variance $\sigma^2 = 1$.

Problem 3: The diffusion approximation 3 points

Back to neuroscience. The diffusion approximation of Stein's model[1] is used to increase the model's tractability. In this approximation, the incoming excitatory and inhibitory jumps are replaced by a Gaussian white noise diffusion term. Summarise why and when (what types of input) this approximation is well founded by making reference the central limit theorem.

Solution.

For Stein's model, we have excitatory and inhibitory inputs being independent Poisson processes with rates λ_E and λ_I, producing jumps J_E and J_I, respectively. The input at each time step is thus a new random variable $I(t) = J_E S_E(t) - J_I S_I(t)$ for Poisson processes $S_E(t)$, $S_I(t)$. From the central limit theorem, we know that adding enough of these contributions (excitatory and inhibitory jumps) together will give this new variable a distribution tending to normal. The conditions, thus, in which the Poisson inputs can be replaced by some Gaussian noise are as follows.

In order to retain temporal resolution in the model, and thus not have to wait too long periods of time for the sum of the inputs to look Gaussian, the rate of inputs $\lambda_E + \lambda_I$ should be large enough. Secondly, so that the input can be modelled as continuous noise rather than composed of discrete jumps (as the Poisson input), the Poisson jumps J_E and J_I should be sufficiently small. If both these conditions are met, the input can be replaced by some continuous Gaussian noise, generally Gaussian white noise if values of the input at different times are independent (as is for a Poisson process, having covariance vanishing apart from zero lag). This allows Stein's LIF model to be replaced by the Ornstein-Uhlenbeck process, where shotnoise is replaced by the Wiener process (the Wiener process being the integral of Gaussian white noise). Here as in the LIF model, the membrane potential is allowed to evolve in time until it reaches threshold whereby a spike is elicited and the potential is reset.

[1] A leaky integrate and fire model with excitatory and inhibitory input producing "jumps" in the membrane potential, see lecture 7.