Modelo neuronal estocástico de FitzHugh-Nagumo. Propiedades pobabilísticas y estimación.

IV Jornadas Estadística Aplicada. La Paloma

J.R. León and A. Samson
IMERL and LJK Grenoble

In this talk we will consider a neuronal model called the FitzHugh-Nagumo (FHN) model, which is a two-dimensional model with noise on the second coordinate and is more simple than other neuronal models. This type of models are variants of the famous Hodgkin-Huxley model established by Hodgkin in 1952.

The stochastic FitzHugh-Nagumo model is defined as follows. Let us denote X_t the membrane potential of the neuron at time t and C_t represents the channel kinetic. We assume that $((X_t, C_t),\ t \geq 0)$ is governed by the following Itô stochastic differential equation (SDE):

$$\begin{cases} dX_t &= \frac{1}{\varepsilon}(X_t - X_t^3 - C_t - s)dt, \\ dC_t &= (\gamma X_t - C_t + \beta)\, dt + \tilde{\sigma} dW_t, \end{cases} \tag{1}$$

where W_t is a standard Brownian motion, parameter s is the magnitude of the stimulus current, $\tilde{\sigma}$ the diffusion coefficient, γ, β are positive constants and ε is a time constant.

Given that in the above system the noise only appears in the second coordinate, the process solution of such a system is a Hypoelliptic diffusion. These systems are well known by physicists and the more simple of them is the linear harmonic oscillator perturbed by a white noise. This process can be defined as the solution of following second order stochastic differential equation

$$X_t'' + \gamma X_t' + \omega_0^2 X_t = \sigma W_t'.$$

In this equation ω_0 is the oscillator own frequency, γ is the friction (damping) coefficient and W is a Brownian motion.

This equation can be written in the form of a system by setting $Y_t = X_t'$ thus

$$\left\{ \begin{array}{rcl} dX_t & = & Y_t dt, \\ dY_t & = & -\left(\gamma Y_t + \omega_0^2 X_t\right) dt + \sigma dW_t. \end{array} \right. \tag{2}$$

This type of system is called a Hamiltonian system. The Hamiltonian is defined as the function $H(x,y) = \frac{1}{2}y^2 + V(x)$ where the potential is the function $V(x) = \omega_0^2 x^2$. In this case we can see that the stochastic system has an invariant measure given by the density

$$p_{inv}(x,y) = \frac{1}{C} e^{-\frac{2\gamma}{\sigma} H(x,y)}.$$

Here C is a normalization constant.

Yielding that the invariant measure is a bi-dimensional Gaussian. Moreover, the process $Z_t = (X_t, Y_t)$ is Markov and β-mixing exponential. These properties provide a frame where is possible to make estimation both parametric and non-parametric. In two (2001) articles written by D. Talay and L. Wu, this frame was extended to include other models sharing the same ergodic properties. The system writes now as

$$\begin{cases} dX_t &= Y_t dt, \\ dY_t &= -\left(c(X_t, Y_t)Y_t + V'(X_t)\right) dt + \sigma(X_t, Y_t) dW_t. \end{cases} \tag{3}$$

The parameters that define the model are assumed to satisfy the following hypothesis. The potential V is lower bounded and continuously differentiable. The damping function c is bounded over compact, and the function $0 < \sigma$ is bounded. The system of FHN can be written in such a form making the following change of variable. More precisely, let set $Y_t = \frac{1}{\varepsilon}(X_t - X_t^3 - C_t - s)$. Then the FitzHugh-Nagumo system (1) can be written as:

$$\begin{cases} dX_t &= Y_t dt, \\ dY_t &= -(c(X_t)\, Y_t + \partial_x V(X_t))dt + \sigma\, dW_t. \end{cases} \tag{4}$$

Where we have the damping force $c(x) = \frac{1}{\varepsilon}(3x^2 - 1 + \varepsilon)$, the potential $V(x) = \frac{x^4}{4} + \frac{\gamma-1}{2}x^2 + (s+\beta)x$ and the diffusion coefficient $\sigma = \frac{\tilde{\sigma}}{\varepsilon} > 0$.
The infinitesimal generator $\mathcal{L}$ of system (4) is

$$\mathcal{L} = \frac{\sigma^2}{2}\,\partial_{yy} + y\,\partial_x - (c(x)\,y + \partial_x V(x))\partial_y\,.$$

System (4) verifies the following conditions:

- The potential $V(x) = \frac{x^4}{4} + \frac{\gamma-1}{2}x^2 + (s+\beta)x$ is lower bounded, smooth over $\mathbb{R}$, V and V' have polynomial growth at infinity.
- The damping coefficient $c(x) = \frac{1}{\varepsilon}(3x^2 - 1 + \varepsilon)$ is continuous, for all $N > 0$, $\sup_{|x|\leq N} |c(x)| < +\infty$ and for all $x \in \mathbb{R}$,

$$c(x) \geq 1 - \frac{1}{\varepsilon} =: c$$

- The diffusion coefficient is $\sigma = \frac{\tilde{\sigma}}{\varepsilon}$. We assume that the (unknown) parameters ε and σ are such that $\varepsilon \in [\varepsilon_0, \varepsilon_1]$ and $\tilde{\sigma} \in [\tilde{\sigma}_0, \tilde{\sigma}_1]$. Then we have $0 < \sigma < \sigma_1 := \frac{\tilde{\sigma}_1}{\varepsilon_0}$.

These three conditions imply that assumptions (H1), (H2) and (H3) of Wu are fulfilled.
We can also notice that the force $-V'(x)$ is "strong enough" for $|x|$ large to make the system return quickly to compact subsets of $\mathbb{R}^2$, which corresponds to condition (0.5) of Wu. Indeed, we have

$$\frac{xV'(x)}{|x|} \to +\infty \,, \quad \text{as } |x| \to +\infty.$$

Let $Z_t = (X_t, Y_t)$ be the solution of system (4). The stochastic process (Z_t) is hypoelliptic and strong Feller. We now state the existence and uniqueness of an invariant probability, and also that the process is β-mixing. It is important to define as in Wu the Lyapounov function. For this let define first the function

$$F(x,y) = a\left(\frac{1}{2}y^2 + V(x)\right) + by\,G(x) + y\,W'(x) + b\,U(x),$$

where G, W and F are three conveniently chosen functions and a and b are constants. Then following Wu, we choose for the Lyapounov function

$$\Psi(x,y) = e^{F(x,y) - \inf_{\mathbb{R}^2} F} \tag{5}$$

Theorem

Let (X_t, Y_t) be solution of system (4).

1. *The process (X_t, Y_t) is positive recurrent with a unique invariant probability measure μ.*
2. *Moments of any order of μ exist: for all $k_1, k_2 \in \mathbb{N}$,* $\mathbb{E}(X_t^{k_1} Y_t^{k_2}) = \int x^{k_1} y^{k_2} d\mu(x, y) < +\infty$.

The process (X_t, Y_t) is mixing. This means that there exist constant $D > 0$ and $0 < \rho < 1$ such that for all z,

$$\left| P_t f(z) - \int f d\mu \right| \leq D \sup_a \left(\frac{\left| f(a) - \int f d\mu \right|}{\Psi(a)} \right) \Psi(z) \, \rho^t \, . \qquad (6)$$

where Ψ is a Lyapounov function defined by (5).

Finally it can be proved that also one has the following inequality in norm of total variation

$$||P_t(z,\cdot) - \mu||_{TV} \leq D''\Psi(z)\rho^t.$$

One can apply this inequality to the skeleton chain $\tilde{Z}_k = Z_{kh}$ for a certain h. This is, if we denotes $\tilde{P}_k$ the discrete semi-group associated to $\tilde{Z}$ we get

$$||\tilde{P}_k(z,\cdot) - \mu||_{TV} \leq D''\Psi(z)\rho^{hk},$$

It is well known that the following equality holds true

$$\beta_h(k) = \int ||\tilde{P}_k(z,\cdot) - \mu||_{VT}\, d\mu(z) \leq D''||\Psi||_1 \rho^{hk}.$$

So we have $\beta_h(k) \leq D''||\Psi||_1 \rho^{hk}$. Hence the skeleton chain is exponentially β-mixing.
Hypoelliptic property of the system ensures that the distribution $P_t(z,\cdot)$ of the process Z_t starting from $Z_0 = z$ has a smooth density $p_t(z,\cdot)$, and the invariant measure has also a smooth density $\mu(dz) = p(z)dz$. In the following, we will denote p^x and p^y the marginal of p with respect to x and y.

In the neuronal context, the ion channel coordinate Y_t can not be measured and only discrete observations of X at discrete times $i\delta$, $i = 1, \ldots, n$ with discretization step δ are available. The density p has no explicit formula. We therefore propose a non-parametric estimation of p from observations $(X_{1\delta}, \ldots X_{n\delta})$. In this section we apply to our context the recent work of Comte, Prieur and Sansom "*Adaptive estimation for Stochastic Damping Hamiltonian Systems under Partial Observation*".

Let K be some kernel $\mathcal{C}^2$ function with compact support A such that its partial derivatives functions $\frac{\partial K}{\partial x}$ and $\frac{\partial K}{\partial y}$ are in $\mathbb{L}^2(\mathbb{R})$, $\int K(x,y)dxdy = 1$ and $\int K^2(x,y)dxdy < \infty$. For all bandwidth $b = (b_1, b_2)$ with $b_1 > 0, b_2 > 0$, for all $(x,y) \in \mathbb{R}^2$, we denote

$$K_b(x,y) = \frac{1}{b_1 b_2} K\left(\frac{x}{b_1}, \frac{y}{b_2}\right).$$

When both coordinates are observed, the natural estimator of p for all $z = (x,y) \in \mathbb{R}^2$, is:

$$\begin{aligned}\tilde{p}_b(z) &= \tilde{p}_b(x,y) \\ &= \frac{1}{n}\sum_{i=1}^{n} K_b\left(x - X_{i\delta}, y - Y_{i\delta}\right) = \frac{1}{n}\sum_{i=1}^{n} K_b(z - Z_{i\delta}).\end{aligned}$$

When only X is observed, we replace Y by increments of X. Indeed, for any $i = 1, \ldots, n$, when δ is small enough, we have

$$X_{(i+1)\delta} - X_{i\delta} = \int_{i\delta}^{(i+1)\delta} Y_t dt \approx \delta Y_{i\delta}$$

Let us thus define the following approximation of $Y_{i\delta}$:

$$\bar{Y}_{i\delta} = \frac{X_{(i+1)\delta} - X_{i\delta}}{\delta}$$

and define the 2-dimensional kernel estimator by

$$\hat{p}_b(x, y) := \frac{1}{n} \sum_{i=1}^{n} K_b \left(x - X_{i\delta}, y - \bar{Y}_{i\delta} \right) \ (*).$$

The bandwidth $b = (b_1, b_2)$ has to be chosen to realize a trade-off between the bias of $\hat{p}_b$ and its variance. This is automatically achieved using the adaptive estimation procedure proposed by the cited work of Comte et al. We can apply their procedure because we have already proved that the invariant density p decreases exponentially and is β-mixing.

Their procedure, inspired by Goldenshluger and Lepski, is the following. Let $\mathcal{B}_n = \{(b_{1,k}, b_{2,\ell}), k, \ell = 1/\sqrt{n}, \dots, c/\sqrt{n}\}$ be the set of possible bandwidths. Set for all $z = (x, y)$ and all $b, b' \in \mathcal{B}_n$

$$\hat{p}_{b,b'}(z) = K_{b'} \star \hat{p}_b(z) = \frac{1}{n} \sum_{i=1}^{n} K_{b'} \star K_b(x - X_{i\delta}, y - \bar{Y}_{i\delta}).$$

Now let

$$A(b) = \sup_{b' \in \mathcal{B}_n} \left(\|\hat{p}_{b,b'} - \hat{p}_{b'}\|^2 - V(b') \right)_+$$

with

$$V(b) = \kappa_1 \frac{1}{nb_1 b_2} \sum_{i=0}^{n-1} \beta(i\delta) + \kappa_2 \frac{\delta}{b_1 b_2^3},$$

where κ_1, κ_2 are numerical constants and $\beta(i\delta)$ are the β-mixing coefficients. The selection is then made by setting

$$\hat{b} = \arg\min_{b \in \mathcal{B}_n} \left(A(b) + V(b)\right) \tag{7}$$

Comte et al. prove an oracle inequality for the final estimator $\hat{p}_{\hat{b}}$.

Theorem (Comte et al's result)

Set $p_b(z) = K_b \star p$ the function that is estimated without bias by $\hat{p}_n$. We have

$$\mathbb{E}(\|\hat{p}_{\hat{b}} - p\|^2) \leq C \inf_{b \in \mathcal{B}_n} \left(\|p_b - p\|^2 + V(b)\right) + C\frac{\log(n)}{n\delta}$$

As explained in Comte et al., the Goldenshluger and Lepski's procedure is numerically demanding due to the double convolutions $\hat{p}_{b,b'}$, especially in the multidimensional case. They therefore propose a simplified procedure based on a work of Lacour, Massart and Rivoirard, that we also implement in this case. The selection of the bandwidth is the following:

$$\hat{\hat{b}} = \arg\min_{b \in \mathcal{B}_n} \left(\|\hat{p}_b - \hat{p}_{b_{min}}\|^2 + V(b) \right) \tag{8}$$

with $\kappa_1 = 0.1$ and $\kappa_2 = 0.001$ and $b_{min} = (1/\sqrt{n}, 1/\sqrt{n})$. By plugging $\hat{\hat{b}}$ into (*) we obtain $\hat{p} := \hat{p}_{\hat{\hat{b}}}$ which is the final estimator of p.

We are going to consider the existing link between the spikes of the process Z_t and its up-crossing of high levels. Let us denote N_t the number of pulses (spikes) during interval time $[0, t]$. Let us define the spike rate as

$$\rho := \lim_{t \to \infty} \frac{N_t}{t}.$$

The spike rate can also be defined as follows. Let us denote T_i the ith interspike interval, i.e. the time between the i and $i + 1$ spikes.

The mean time $< T >$ between two spikes can be defined using the ergodic theorem as

$$< T >= \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} T_i .$$

As pointed out by Linder & Schimansky-Geier there exists a relationship between these two measures. In fact

$$\rho = \lim_{N\to\infty} \left(\frac{T_0}{N} + \frac{1}{N} \sum_{i=1}^{N} T_i + \frac{T_{N+1}}{N} \right)^{-1} = \frac{1}{< T >},$$

where T_0 and T_{N+1} are the time intervals until the first or after the last spike.

We know that process $Z_t = (X_t, Y_t)$ defines a mesure $\mathbb{P}^z$ in the space $\Omega := C(\mathbb{R}^+, \mathbb{R}^2)$. This means that X_t is an a.s. continuously differentiable process and $\dot{X}_t = Y_t$. Let us define the number of up-crossings of process $X.$ at level u in $[0, t]$ as
$U_t^{X.}(u) = \#\{s \leq t : \ X_s = u, \ Y_s > 0\}$.
Now we want to link ρ with the up-crossing process. If we forget the boundary effects, the random variable N_t will be equal to the number of up-crossings at level u, for u large enough. Hence using this approximation we finally obtain

$$\rho = \frac{1}{< T >} \approx \rho(u) = \lim_{t \to \infty} \frac{U_t^{X.}(u)}{t} \tag{9}$$

In (9) two things need some formal clarification. The first one is the existence of the a.s. limit. The second one corresponds to the approximation.

For the first one firstly one can prove the following Rice's formula on the expectation of the number of up-crossings.

Theorem

Let $Z_t = (X_t, Y_t)$ be the stationary solution of the FitzHugh-Nagumo system. The Rice's formula holds true

$$\mathbb{E}U_t^{X_\cdot}(u) = t\int_0^\infty yp(u,y)dy. \tag{10}$$

Moreover the ergodic theorem implies that

$$\lim_{t \to \infty} \frac{U_t^{X_\cdot}(u)}{t} = \int_0^\infty y p(u, y) dy = \rho(u), \; a.s.$$

The second point, the approximation, is more involved and was settled down in the Cramer-Leadbetter's (CL) book.
We will sketch the procedure adapt to our case.

We know by using the ergodic theorem that

$$U^{X.}_{[0,t]}(u) = t\rho(u) + o(t).$$

Let us denote $C^{X.}_{[0,t]}(u)$ the number of all crossings on the interval $[0, t]$. And let us define the sets

$$H_1(\tau, t) = \mathbb{P}\{U^{X.}_{(-\tau,0)}(u) \geq 1,\ C^{X.}_{(0,t)}(u) \leq 1\}.$$

In CL book it is shown that

$$\lim_{\tau \to 0} \frac{H_1(\tau, t)}{U^{X.}_{(0,\tau)}(u)} = \Phi_1(t)$$

for a finite function Φ_1.

This last function represents the conditional probability of no more than 1 crossings in the interval $(0, t)$, given that an upcrossing occurred "at" time zero. Now the important function for us is

$$F_2(t) = 1 - \Phi_1(t).$$

The function F_2 may be regarded as the "distribution function of the length of the interval of an arbitrarily chosen upcrossing and the next upcrossing".

Introducing now the function $u_0(t) = \mathbb{P}\{U^{X_\cdot}_{(0,t)} = 0\}$. In CL it is proved that

$$\lim_{\tau^+ \to 0} \frac{u_0(t+\tau) - u_0(t)}{\tau} = -\rho(u)\Phi_1(t).$$

Thus the function u_0 has right-hand side derivative, in fact $D^+ u_0(t) = -\rho(u)\Phi_1(t)$.

Let us recall some properties of function F_2. It holds $0 \leq F_2(t) \leq 1$. Moreover the Lebesgue theorem gives

$$u_0(T) - u_0(0) = \int_0^T D^+ u_0(t) dt,$$

but as u_0 is bounded, the derivative $D^+ u_0(t)$ is integrable over $(0, \infty)$. This in particular implies that $D^+ u_0(t) \to 0$ whenever $t \to \infty$, thus $\lim_{t\to\infty} \Phi_1(t) = 0$ and therefore $F_2(t) \to 1$ when $t \to \infty$. Finally as F_2 is non-decreasing it is a real distribution function.

Let us compute now the mean of this distribution function.

$$\int_0^\infty t dF_2(t) = \int_0^\infty [1 - F_2(t)]dt = \frac{1}{\rho(u)}[u_0(0) - u_0(\infty)] = \frac{1}{\rho(u)},$$

the last equality comes from that $u_0(0) = 1$ and $u_0(\infty) = 0$.
Given a justification to the approximation procedure.
A formula also holds for the second moment this is

$$\int_0^\infty t^2 dF_2(t) = \frac{2}{\rho(u)} \int_0^\infty u_0(t)dt.$$

However, the function u_0 is not known in our model, but we also built an estimator.

Equation (9) provides a good start to estimate the spike rate. The quantity that we estimate is not directly ρ but rather $\rho(u)$ for a large level u. Using $\hat{p}$ the kernel estimator of the invariant density, we define the following estimator of $\rho(u)$:

$$\hat{\rho}(u) = \int_0^\infty y\hat{p}(u,y)\,dy.$$

For instance let us consider a multiplicative two-dimensional kernel $K(x,y) = k(x)k(y)$, where k is a continuous and bounded kernel, such that $\int k(v)dv = 1$. Then we have

$$\hat{p}(u,y) = \frac{1}{nb_1b_2}\sum_{i=1}^{n} K\left(\frac{u - X_{i\delta}}{b_1}, \frac{y - \bar{Y}_{i\delta}}{b_2}\right).$$

Set $\tilde{k}_1 = \int_0^\infty yk(y)dy$.

We get

$$\hat{\rho}(u) = b_2 \tilde{k}_1 \hat{p}^x(u) + \frac{1}{2nb_1} \sum_{i=1}^{n} k\left(\frac{u - X_{i\delta}}{b_1}\right) \bar{Y}_{i\delta},$$

here $\hat{p}^x(\cdot)$ is the kernel estimator of the first marginal density of the invariant measure. For a gaussian centered kernel k, we obtain:

$$\hat{\rho}(u) \;=\; \frac{b_2}{2} \hat{p}^x(u) + \frac{1}{2} \frac{1}{nb_1} \sum_{i=1}^{n} k\left(\frac{u - X_{i\delta}}{b_1}\right) \bar{Y}_{i\delta},$$

Furthermore, a CLT can be concluded for $\hat{\rho}(u)$.

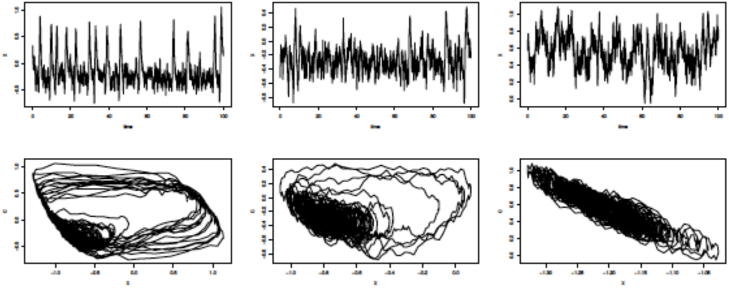

FIG 1. *Simulations of the FitzHugh-Nagumo. Voltage variable X_t versus time (top line) and the corresponding trajectory in phase space C_t versus X_t (bottom line) for $s = 0$, $\beta = 0.8$, $\tilde{\sigma} = 0.3$, left: $\varepsilon = 0.1$ and $\gamma = 1.5$, middle: $\varepsilon = 0.4$ and $\gamma = 1.5$, right: $\varepsilon = 0.5$ and $\gamma = 0.2$.*

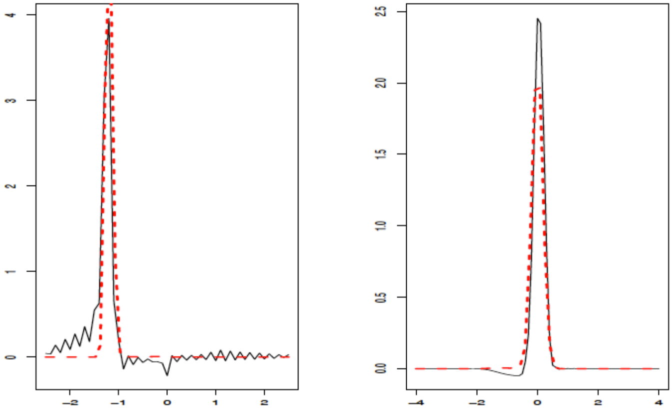

Figure: 1 Estimation of p without spikes

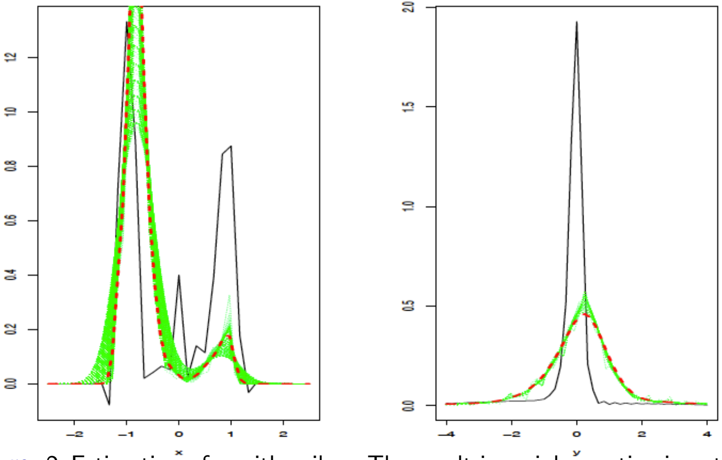

Figure: 2 Estimation of p with spikes. The result is mainly erratic given the non-stable code for the solution of the Langevin PDE.

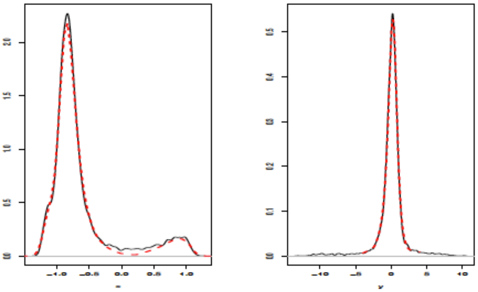

FIG 4. *Invariant density estimation. Left: marginal in x of the estimation $\hat{p}_b$ (red line) and true density approximated by a Monte Carlo scheme (black line). Right: marginal in y of the estimation $\hat{p}_b$ (red line) and true density approximated by a Monte Carlo scheme (black line). Simulations are performed with parameters that generate spikes ($s = 0$, $\beta = 0.8$, $\tilde{\sigma} = 0.3$, $\varepsilon = 0.1$ and $\gamma = 1.5$).*

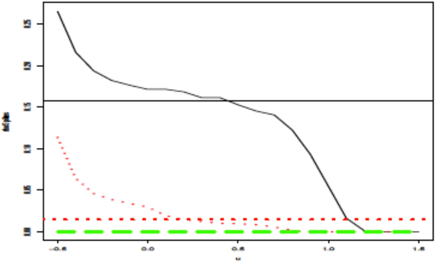

FIG 5. *Spike rate estimators $\hat{\lambda}(u)$ and $\hat{\rho}$ computed as the mean of $\rho(0.1), \ldots, \rho(0.6)$. Black plain curve: evolution of $\hat{\lambda}(u)$ with u and black plain line $\hat{\rho}$ for a set of parameters that generate spikes ($s = 0$, $\beta = 0.8$, $\tilde{\sigma} = 0.3$, $\varepsilon = 0.1$ and $\gamma = 1.5$). Red dotted curve: evolution of $\hat{\lambda}(u)$ with u and red dotted line $\hat{\rho}$ for a set of parameters that generates few and small excursions ($s = 0$, $\beta = 0.8$, $\tilde{\sigma} = 0.3$, $\varepsilon = 0.4$ and $\gamma = 1.5$). Green dashed curve: evolution of $\hat{\lambda}(u)$ with u for a set of parameters that does not generate spikes ($s = 0$, $\beta = 0.8$, $\tilde{\sigma} = 0.3$, $\varepsilon = 0.5$ and $\gamma = 0.2$).*