Bats Monitoring: A Classification Procedure of Bats Behaviors based on Hawkes Processes

Rencontres de la Chaire MMB 2024

Romain Lacoste, PhD student
Supervised by C. Denis, C. Dion-Blanc, L. Sansonnet and Y. Bas

Veolia

October 09, 2024

Table of contents

1. Ecological and mathematical setting
2. Statistical methodology
3. Numerical experiments
4. Application on real data

Ecological problematic and motivation

Two behaviors:

- **commuting** mode;
- **foraging** mode.

Goal: predicting the majority behavior of bats at sites throughout France.

- discriminate the **foraging** behavior from the **commuting** behavior.

Motivations:

- contribute to address spatial ecology issues;
- automate decision-making with few input variables.

Data: time of echolocation calls of **differents species** recorded as part of **Vigie-Chiro** participatory project.

- we focus on the **Common Pipistrelle**.

VIGIECHIRO

Echolocation: used by bats for **foraging** and **commuting**.

Behavioral characterization: via the way bats emit calls (see Griffin *et al.* (1960)).

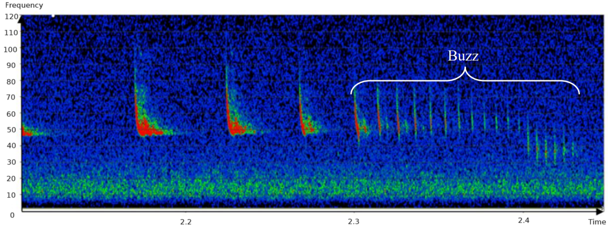


Figure: Sonogram containing a feeding buzz.

- consider the temporal distribution of the calls.
- sequence of calls $(T_\ell)_{\ell \geq 1}$ as a realization of a point process N.

Modeling the sequence of calls

Point processes: model the occurrence of random events over time.

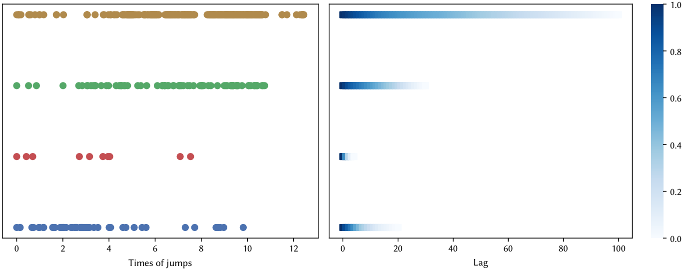


Figure: Left: the start times of echolocation calls sequences, right: autocorrelation as a function of the lag for four nights.

- presence of strong temporal dependence in data.

Data modeling with Hawkes processes

The linear exponential Hawkes process : a point process N with conditional intensity function (see Daley and Vere-Jones (2003)):

$$\lambda_\theta(t) := \mu + \int_0^t \alpha\beta e^{-\beta(t-s)} dN_s = \mu + \sum_{T_\ell < t} \alpha\beta e^{-\beta(t-T_\ell)},$$

where :
- $(T_\ell)_{\ell \geq 1}$ the **time jumps** of the process;
- $\theta \in \Theta = \{\mu > 0, 0 \leq \alpha < 1, \ \beta \geq 0\}$
- $\mu \longrightarrow$ **exogenous intensity**;
- $\alpha \longrightarrow$ **arrival intensity**;
- $\beta \longrightarrow$ **rate of the decay**.

Modelisation: **the start time** of a call correspond to a **jump** of the Hawkes process.

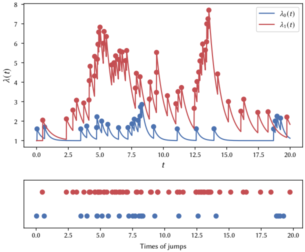
Classification model

Let $\mathcal{D}_n^L = \left\{(\mathcal{T}_T^1, Y^1), \ldots, (\mathcal{T}_T^n, Y^n)\right\}$ be a sample of i.i.d. observations such that:

- **Label:** $Y \sim \mathcal{B}\left(p^*\right), \; Y \in \{0, 1\}$;
- **Feature:** $\mathcal{T}_T = \left(T_1, \ldots, T_{N_T}\right)$ of intensity $\lambda_{\theta^*_Y}(t)$ on $[0, T]$ with $\theta^*_Y \in \Theta$.

Goal: learn a decision rule g from $\mathcal{D}_n^L$ such that $g(\mathcal{T}_T)$ is a prediction of the label Y.

▶ given a new unlabeled feature $\mathcal{T}_T^{n+1}$, our guess for Y^{n+1} is $g\left(\mathcal{T}_T^{n+1}\right)$.



Quality of label prediction: measured by its missclassification risk

$$\mathcal{R}(g) := \mathbb{P}\left(g\left(\mathcal{T}_T^{n+1}\right) \neq Y^{n+1}\right).$$

Bayes rule and empirical risk minimization

Bayes rule: characterized by

$$g_{p^*,\boldsymbol{\theta}^*}(\mathcal{T}_T) = \mathbb{1}_{\left\{\eta_{p^*,\boldsymbol{\theta}^*}(\mathcal{T}_T) > \frac{1}{2}\right\}}$$

where $\eta_{p^*,\boldsymbol{\theta}^*}(\mathcal{T}_T) := \mathbb{P}(Y = 1 | \mathcal{T}_T) = \dfrac{p^* \exp\Big(F_{\theta_1^*}(\mathcal{T}_T)\Big)}{p^* \exp\Big(F_{\theta_1^*}(\mathcal{T}_T)\Big) + (1-p^*) \exp\Big(F_{\theta_0^*}(\mathcal{T}_T)\Big)}$

Empirical risk: based on $\mathcal{D}_n$ estimates $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{Y^i=1\}}$ and solve :

$$\hat{\boldsymbol{\theta}} \in \underset{\boldsymbol{\theta}\in\Theta^2}{\operatorname{argmin}} \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{g_{\hat{p},\boldsymbol{\theta}}(\mathcal{T}_T^i) \neq Y^i\}}$$

- minimize this require to solve a non convex optimization problem.

ERM procedure

Convexification: replace the $0-1$ loss by a **convex surrogate** (see Zhang (2004)) and based on $\mathcal{D}_n$ solve instead :

$$\hat{\theta} \in \underset{\theta \in \Theta^2}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} \left(Z^i - f_{\hat{p},\theta}(\mathcal{T}_T^i) \right)^2$$

where $Z^i = 2Y_i - 1$ and $f_{\hat{p},\theta}(\mathcal{T}_T) = 2\eta_{\hat{p},\theta}(\mathcal{T}_T) - 1$ with

$$\eta_{\hat{p},\theta}\left(\mathcal{T}_T\right) = \frac{\hat{p} \exp\left(F_{\theta_1}(\mathcal{T}_T)\right)}{\hat{p} \exp\left(F_{\theta_1}(\mathcal{T}_T)\right) + (1-\hat{p}) \exp\left(F_{\theta_0}(\mathcal{T}_T)\right)}$$

Classifier: $\hat{g}(\mathcal{T}_T) = \mathbb{1}_{\left\{\hat{f}(\mathcal{T}_T) \geq 0\right\}}$.

ERM procedure: provides estimates of (θ_0^*, θ_1^*).

- gives a model for the behavior within each class.

Model evaluation: by performing a goodness-of-fit test.

- using the **Time-Rescaling Theorem** (see Daley and Vere-Jones (2003)):

Theorem

Let $\Lambda(t) = \int_0^t \lambda(s)\, ds$ be the **compensator** of the process N. Then, a.s., the transformed sequence $\{\tau_j = \Lambda(T_j)\}$ is a realization of a unit-rate Poisson process if and only if the original sequence $\{T_j\}$ is a realization from the point process N.

Test H_0: "*the sequence of observations is a realization of the point process with intensity $\lambda_{\hat{\theta}_k}$*".

- test if $\{\Lambda_{\hat{\theta}_k}(T_{j+1}) - \Lambda_{\hat{\theta}_k}(T_j)\} \overset{\text{iid}}{\sim} \mathcal{E}(1)$

Simulation: by cluster representation using the branching properties of the self-exciting Hawkes process (see Hawkes and Oakes (1974)).

Panel of scenarios:

	Scenario 1	*Scenario 2*
μ_0	1.0	1.0
μ_1	1.0	0.5
α_0	0.2	0.0
α_1	0.7	0.5
β_0	3.0	0.0
β_1	1.5	1.5

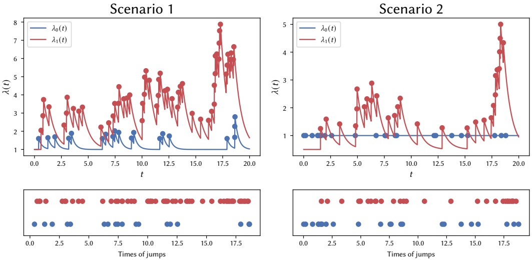


Table: Scenario panel used to study procedure performance.

Simulation scheme: 20 Monte-Carlo repetitions in each scenario.

In both scenarios: $n_{\text{train}} = 300$, $n_{\text{test}} = 1000$, $T = 20$, $p^* = 0.5$.

Empirical error rate: $\hat{\mathcal{R}}_n(g) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\left\{g(\mathcal{T}_T^{(i)}) \neq Y_i\right\}}$.

	Error Rate		
	Bayes	ERM	RF
Scenario 1	0.07 (0.00)	0.07 (0.00)	0.09 (0.01)
Scenario 2	0.17 (0.00)	0.17 (0.01)	0.30 (0.03)

Table: Empirical error averaged over 20 repetitions.

Goodness-of-fit test: if $\hat{g}\left(\mathcal{T}^i\right) = k$ test if $\{\Lambda_{\hat{\theta}_k}(T^i_{j+1}) - \Lambda_{\hat{\theta}_k}(T^i_j)\} \overset{\text{iid}}{\sim} \mathcal{E}(1)$

		$\hat{g}(\mathcal{T})$	
		p-value	Acceptance Rate
Scenario 1	Class 0	0.51 (0.01)	0.96 (0.01)
	Class 1	0.51 (0.03)	0.95 (0.02)
Scenario 2	Class 0	0.41 (0.01)	0.89 (0.01)
	Class 1	0.41 (0.03)	0.90 (0.01)

Table: Mean p-values and acceptance rate for a 5% significance level test over 20 repetitions.

Real data

- Calls recorded over one night at 755 sites in France.

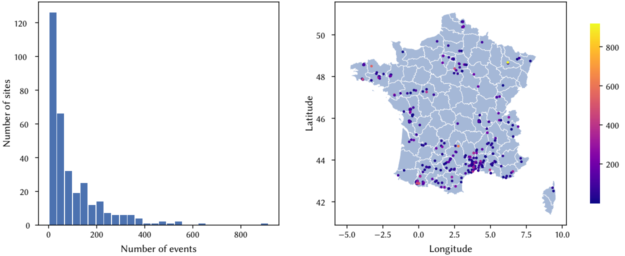


Figure: Each point on the map represents a site and its colour refers to the number of events in the temporal sequences.

- 332 labeled sites.
- 423 unlabeled sites.

Classification on labeled data

Assess the performance: by comparing with labels given by the metric.

Evaluation scheme: repeat 20 times:

- choose 75% for training and the remaining 25% for testing.

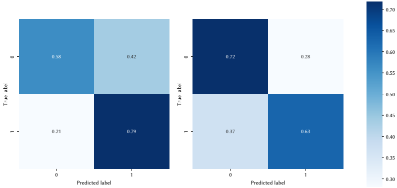


Figure: Confusion matrix of prediction on $\mathcal{D}^L_{n_{\text{test}}}$. Score: ERM: 68.13% (4.15), RF: 67.35% (2.21).

	$\hat{g}(\mathcal{T})$	
	p-value	Acceptance Rate
Class 0	0.26 (0.06)	0.66 (0.11)
Class 1	0.15 (0.03)	0.45 (0.07)

Table: Mean p-values and reject rate for a 5% significance level test on $\mathcal{D}^L_{n_{\text{test}}}$.

Prediction on intermediate sites: tricky since bats have mixed behavior.

Training: based on labeled data. ► $(\hat{\theta}_0, \hat{\theta}_1)$, $\hat{\eta}$, $\hat{g}$.

Goodness-of-fit test: on unlabeled data

	$\hat{g}(\mathcal{T})$	
	p-value	Acceptance Rate
Class 0	0.15	0.43
Class 1	0.21	0.49

Table: Mean p-values and acceptance rate for a 5% significance level test.

Discussion:

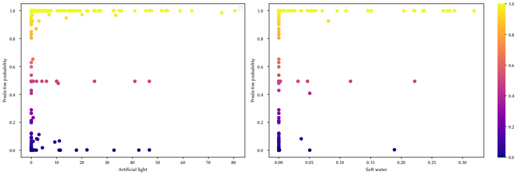


Figure: Predictive probability given by $\hat{g}$ on $\mathcal{D}_n^U$ as a function of environmental covariates.

Conclusion:

- validation of the procedure on synthetic data;
- Hawkes processes modeling: revelant for echolocation calls data;
- classification procedure: prediction and behavioral confidence index;
- provides a tool to ecologist for predicting bats behavior.

Bats Monitoring: A Classification Procedure of Bats Behaviors based on Hawkes Processes, C. Denis, C. Dion-Blanc, R.E. Lacoste, L. Sansonnet and Y. Bas (2023), The Journal of the Royal Statistical Society, Series C.

Ongoing work (1/2)

Future exploration: consider species with more marked majority behavior.

- data processing for the **Western Barbastelle**, **Daubenton's myotis**.

Model enrichment: by considering multivariate Hawkes processes.

- model **simultaneously** the call sequence of multiple species.
- incorporates the effects of cooperation and competition between species.

ERM-Lasso classification rule for Multivariate Hawkes Processes paths, C. Denis, C. Dion-Blanc, R.E. Lacoste and L. Sansonnet, HAL/arXiv.

- include inhibition interaction in the model.

Ongoing work (2/2)

Package development: user-friendly tools for simulation, estimation and classification of exponential multivariate Hawkes processes.

Features :

- learning for short-time path repetition data.
- suitable for large-scale networks (Lasso procedure).
- code source implemented in C++ for rapid computation.

Sparkle: a statistical learning toolkit for Hawkes process modeling in `Python`, R.E. Lacoste, soon on HAL/arXiv.

- ▶ Available soon on GitHub at **https://github.com/romain-e-lacoste**

Thank you for your attention!

Any questions?