Post-Doctoral position
Laboratory I3S, Sophia Antipolis, France

Title	Identification of a Cox Model for Estimation of Risk of Decompression Sickness
Contact	Luc Pronzato (`pronzato@i3s.unice.fr`) João Rendas (`rendas@i3s.unice.fr`)
Duration	12 months — available to candidates of all nationalities
Salary	about € 2110/month (netto)

Objective The objective of this study is to establish a model for the risk of accidents of decompression sickness arising in the context of underwater diving.

Let the binary random variable $DCS(X) \in \{0, 1\}$ represent the occurrence of an accident (DCS $\equiv$ decompression sickness) during dive X, with X characterizing in particular the depth-duration profile. Our goal is to *predict the risk of accident* $y(X)$ during that dive, with

$$X \longrightarrow y(X) = \text{Prob}\left\{DCS(X) = 1\right\} \in [0,\, 1]\,.$$

Besides predicting this risk we will need to *assess its confidence*, ideally in a form of a confidence interval.

Learning data and existing knowledge

Data The model will be learned from existing data sets gathering, for a large number M of distinct diving profiles:

- the bottom depth, dive duration and mixed gas breathed,
- the decompression procedure followed,
- the numbers of dives n_i and of occurrences a_i of DCS, $i = 1, \ldots, M$.

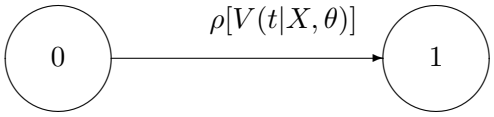
Jointly, the first two items above specify X_i for each of the M dives, while the last gives information about $\{y(X_i)\}_{i=1}^{M}$.

Model A dynamical model for the production of gas bubbles in the diver's body during the decompression phase of the dive — widely believed to be the cause of occurrence of DCS — is available. Denote by $V(t|X, \theta)$ the output of the model at time t for dive X, where θ is a set of biological parameters characteristic of each diver.

$$X \longrightarrow \boxed{\text{Model}} \longrightarrow V(t|X,\theta) \qquad (1)$$
$$\uparrow$$
$$\text{parameters } \theta$$

Previous work done in the project has identified a probability distribution (for a given population of interest) for θ, $\theta \propto p(\mathrm{d}\theta)$, inducing a random process structure in $V(t)$.

Approach Our approach is based on the definition of a *hazard rate* $\rho(V)$ over the phase space of the biophysical model (1), which yields a family of binary Markov random processes $DCS(t|X,\theta) \in \{0,1\}$ that always start at $DCS(0|X,\theta) = 0$, state 1 (indicating the presence of DCS symptoms) being an *absorbing state*:



The parameter vector θ being itself random, $DCS(t|X,\theta)$ is in fact a *Cox process*, i.e., a Poisson process with random intensity function. The times of accident $t_{DCS}(X,\theta)$ are the instants of transition to 1 of $DCS(t|X,\theta)$.

The probability of a DCS accident during dive X, assuming a given hazard rate field $\rho(\cdot)$, is then

$$y(X|\rho) = \mathrm{Prob}\left\{DCS(X) = 1|\rho\right\} = 1 - \int_\Theta \exp\left\{-M_\rho(\infty|X,\theta)\right\} p(\mathrm{d}\theta)\,,$$

where $M_\rho(T|X,\theta)$ is the *cumulative hazard function*

$$M_\rho(T|X,\theta) = \int_0^T \rho\left[V(t|X,\theta)\right] \mathrm{d}t\,.$$

This gives a parametrization of the risk of the different dives in terms of $\rho(\cdot)$, a function defined over the output space of the biophysical model.

Data likelihood/estimation criterion We will consider both classical and Bayesian methods for estimating $\rho(\cdot)$. The hazard rate must be non-negative: $\rho(V) \geq 0$ for all V. Several approaches can be used to deal with this constraint:

1. One may model $z(V) = \log \rho(\cdot)$. This approach is commonly used to impose positivity of estimated functions, see for instance [2]. In the context of Bayesian estimation, where $Z(V)$ is modeled as a realization of a Gaussian process, this leads to the notion of log-Gaussian Cox processes, see [3]. Several authors used non-parametric approaches to identify these models, see [4, 5].

2. One may add a penalty term to the estimation criterion, for instance $P(\rho) = \alpha \int_{\mathcal{V}} \mathbf{1}\{\rho(V) < 0\}\, dV$ where $\mathbf{1}\{A\}$ is the indicator of set A.

3. One may impose constraints directly on the predictor via a suitable choice of observation weights, see, e.g., [6].

The problem at hand in this study presents an additional difficulty, that is related to the fact that the point process itself is not observed, but only the number of realizations (at most one for the pure death process with a single individual) along a one-dimensional curve. Our goal is thus to perform tomography in the state space, with a complex observer whose geometry is induced by the biophysical model. This is expected to raise important problems of observability that should be overcome by a convenient regularization of the hazard rate.

Required expertise The candidate should have a strong background in statistics (candidates familiar with Point Process models and/or non-parametric statistics are preferred) and have a past experience of use of numeric intensive simulation and optimization methods (MCMC, simulated annealing...).

All software will be developed under Matlab; the candidate is expected to be proficient in this mathematical programming language.

References

[1] D.R. Cox and V. Isham (1980). *Point Processes*. Chapman and Hall, London. MR0598033

[2] J.O. Ramsay and B.W. Silverman (2006). *Functional Data Analysis*. Springer, New York.

[3] J. Moller, A.R. Syversveen and R.P. Waagepetersen (1998). Log Gaussian Cox processes. *Scandinavian Journal of Statistics*, Vol. 25, No. 3, pages 451–482.

[4] R.P. Adams, I. Murray, and D.J.C. MacKay (2009). Tractable nonparametric Bayesian inference in Poisson processes with Gaussian process intensities. In *Proc. of the 26th Int. Conf. on Machine Learning*.

[5] T. Zhang and S.C. Kou (2010). Nonparametric inference of doubly stochastic Poisson process data via the kernel method. *The Annals of Applied Statistics*, Vol. 4, No. 4, pages 1913–1941.

[6] P. Hall and L.-S. Huang (2001). Nonparametric kernel regression subject to monotonicity constraints. *The Annals of Stat.*, Vol. 29, No. 3, pages 624–647.