

Caractérisation de la diversité d'une population à partir de mesures quantifiées d'un modèle non-linéaire. Application à la plongée hyperbare

Résumé

Cette thèse propose une nouvelle méthode pour l'estimation non-paramétrique de densité à partir de données censurées par des régions de formes quelconques, éléments de partitions du domaine paramétrique. Ce travail a été motivé par le besoin d'estimer la distribution des paramètres d'un modèle biophysique de décompression afin d'être capable de prédire un risque d'accident. Dans ce contexte, les observations (grades de plongées) correspondent au comptage quantifié du nombre de bulles circulant dans le sang pour un ensemble de plongeurs ayant exploré différents profils de plongées (profondeur, durée), le modèle biophysique permettant de prédire le volume de gaz dégagé pour un profil de plongée donné et un plongeur de paramètres biophysiques connus.

Dans un premier temps, nous mettons en évidence les limitations de l'estimation classique de densité au sens du maximum de vraisemblance non-paramétrique. Nous proposons plusieurs méthodes permettant de calculer cet estimateur et montrons qu'il présente plusieurs anomalies : en particulier, il concentre la masse de probabilité dans quelques régions seulement, ce qui le rend inadapté à la description d'une population naturelle.

Nous proposons ensuite une nouvelle approche reposant à la fois sur le principe du maximum d'entropie, afin d'assurer une régularité convenable de la solution, et mettant en jeu le critère du maximum de vraisemblance, ce qui garantit une forte attache aux données. Il s'agit de rechercher la loi d'entropie maximale dont l'écart maximal aux observations (fréquences de grades observées) est fixé de façon à maximiser la vraisemblance des données. Nous illustrons par divers exemples que la solution proposée offre de meilleures performances, notamment en généralisation, que l'estimateur non-paramétrique classique du maximum de vraisemblance.

Abstract

This thesis proposes a new method for nonparametric density estimation from censored data, where the censoring regions can have arbitrary shape and are elements of partitions of the parametric domain. This study has been motivated by the need for estimating the distribution of the parameters of a biophysical model of decompression, in order to be able to predict the risk of decompression sickness. In this context, the observations correspond to quantified counts of bubbles circulating in the blood of a set of divers having explored a variety of diving profiles (depth, duration) ; the biophysical model predicts of the gaz volume produced along a given diving profile for a diver with known biophysical parameters.

In a first step, we point out the limitations of the classical nonparametric maximum-likelihood estimator. We propose several methods for its calculation and show that it suffers from several problems : in particular, it concentrates the probability mass in a few regions only, which makes it inappropriate to the description of a natural population.

We then propose a new approach relying both on the maximum-entropy principle, in order to ensure a convenient regularity of the solution, and resorting to the maximum-likelihood criterion, to guarantee a good fit to the data. It consists in searching for the probability law with maximum entropy whose maximum deviation from empirical averages is set by maximizing the data likelihood. Several examples illustrate the superiority of our solution compared to the classic nonparametric maximum-likelihood estimator, in particular concerning generalisation performance.