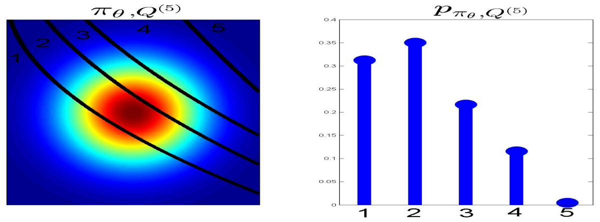
# Most Likely Maximum Entropy for Population Analysis: a case study in decompression sickness prevention

Y. Bennani, L. Pronzato, J. Rendas
Université Nice Sophia Antipolis, CNRS, I3S

## Problem Formulation

- $\Theta \subset \mathbb{R}^d$ space of parameters $\theta$, $\pi_\theta$ probability distribution over $\Theta$,
- $Q$ a $L$-partition of $\Theta$ : $\Theta = \cup_{\ell=1}^{L} Q_\ell, \qquad Q_i \cap Q_j = \emptyset, i \neq j,$
- $p_{\pi_\theta,Q}$ the probability law induced by $\pi_\theta$ on $Q$ : $p_{\pi_\theta,Q_\ell} = \int_{Q_\ell} d\pi_\theta.$




$\tilde{p}_{Q^{(n)}}$ empirical estimate of $p_{\pi_\theta,Q^{(n)}}$, $\nu_n = \frac{N_n}{\sum_{n'=1}^{N} N_{n'}}$.

- Data : samples from distributions laws $\{p_{\pi_\theta,Q^{(n)}}\}_{n=1}^{N}$ determined by $N$ distinct $L$-partitions : $G_i^n \sim p_{\pi_\theta,Q^{(n)}}, i = 1,\dots,N_n, n = 1,\dots,N, \quad G_i^n \in \{1,\dots,L\}$.
- Goal : Determination of $\pi_\theta$ from the $N$ observed samples.
- Log-likelihood :

$$\mathscr{L}(\pi_\theta) = -\sum_{n=1}^{N} \nu_n \left[ \mathcal{D}_{KL}\left(\tilde{p}_{Q^{(n)}} \| p_{\pi_\theta,Q^{(n)}}\right) + \sum_{n=1}^{N} H\left(\tilde{p}_{Q^{(n)}}\right) \right]$$

$\mathcal{D}_{KL}$ : Kullback-Leibler divergence; $H$ : Shanon entropy.

## NPMLE

- support of $\hat{\pi}_\theta$ is confined to $\mathcal{P}$, ($|\mathcal{P}| = K$) a subset of the intersections of the elements of $\{Q^{(n)}\}_{n=1}^{N}$ elements,
- only the probability mass $\mathbf{w}$ of the sets in $\mathcal{P}$ can be estimated.
- non uniqueness the solution

$$\underset{\mathbf{w}}{\text{argmax}} \sum_{n=1}^{N} \sum_{\ell=1}^{L} \tilde{p}_{Q_\ell^{(n)}} \log \mathbf{B}_{\ell.}^{(n)} \mathbf{w}$$

$\mathbf{B}_{\ell.}^{(n)}$ the $\ell$-th row of a $(L \times K)$ binary matrix describing the link between observations $G_i^n$ and elements of $\mathcal{P}$.

- optimization very fast with the modification of Vertex Exchange Method (Harmann et al, Statistics & Probability Letters(2007)).

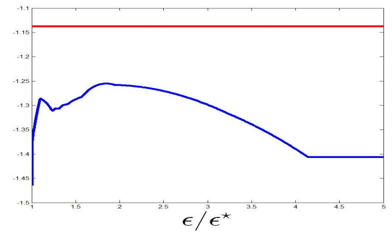
## Most Likely Maxent

- In the large sample where $(N_n \to \infty)$ we should have $\tilde{p}_{Q^{(n)}} \to p_{\pi_\theta,Q^{(n)}}$
- Dudik et al. (Journal of Machine Learning Research, 2007) propose a generalized maxent estimator of the form : $\underset{\pi \in \mathscr{P}}{\text{argmin}} [\mathcal{D}_{KL}(\pi \| q_0) + U(\text{E}_\pi[\tilde{p}_Q])]$ with specific choice of $U$ to express $\ell_1$ or $\ell_2$-regularization, and where features $\tilde{p}_Q$ are obtained from the same empirical distribution.
- We propose to estimate $\pi_\theta$ as the Rényi-maxent (quadratic programming) that best matches observed frequencies : $\epsilon^\star \geq 0$ the smallest value of $\epsilon$ for which there exists a solution to

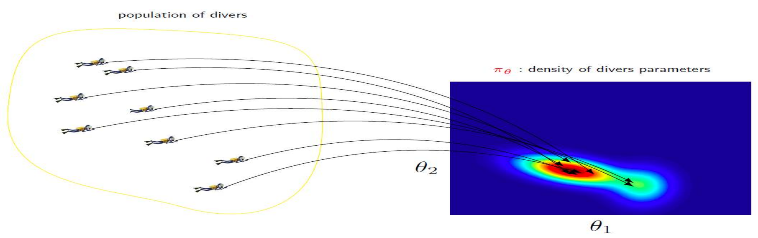
$$\tilde{\pi}_\theta^\epsilon = \underset{\pi}{\text{argmax}} \;\; h_2(\pi) \;\text{ s.t. }\; \left\| \Sigma^{(n)^{-1/2}} \left( p_{\pi_\theta,Q^{(n)}} - \tilde{p}_{Q^{(n)}} \right) \right\|_\infty \leq \epsilon, \;\; \forall n,$$

$h_2(\cdot)$ : Rényi-entropy of order 2; $\Sigma^{(n)}$ : $L \times L$ covariance matrix of the empirical distribution of $\tilde{p}_{Q^{(n)}}$.

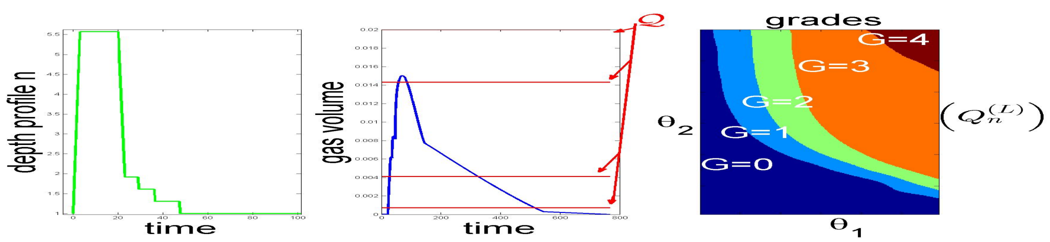
- We propose to use likelihood to chose the degree of regularization $\tilde{\pi}_\theta^{ml} = \underset{\tilde{\pi}_\theta^\epsilon,\, \epsilon > \epsilon^\star}{\text{argmax}} \mathscr{L}(\tilde{\pi}_\theta^\epsilon)$



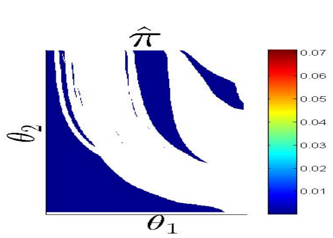

## Application : Decompression sickness



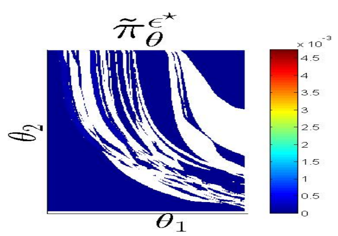

- Accidents due to formation of gas bubbles due to a fast decrease in diver's depth (fast reduction of pressure)
- Only the maximum gas volume $B(\theta, P_i)$ is measured through a (known) quantification $\mathcal{Q} \to$ grades
- From observations of grades $G_i$ we only know that $\theta_i \in \{\theta : \mathcal{Q}(B(\theta, P_i)) = G_i\}$.



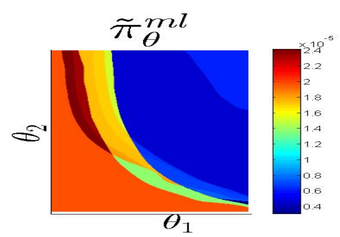

## Results



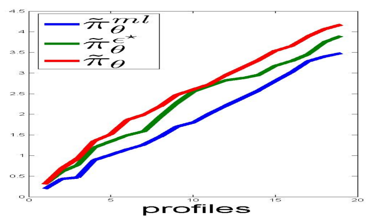

unstable solution not physically plausible biological model of a population

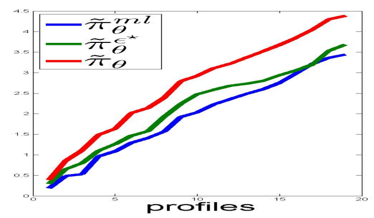


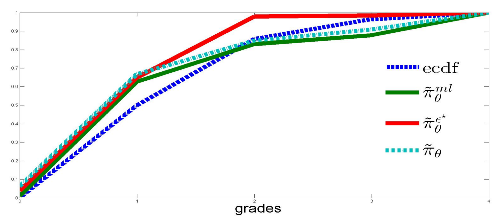

has larger support than $\hat{\pi}_\theta$, still large regions with zero mass




smoother, more plausible and stable, better for prediction




- Leave-one-out-cross-validation, compare kolmogorov and total variation distances between estimated and observed frequencies.
- $\pi_\theta^{ml}$ the most efficient estimator in terms of prediction.