**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 $\Theta$ : $\Theta = \bigsqcup_{i=1}^{L} Q_i$, $Q_i \cap Q_j = \emptyset$, $i \neq j$,
- $p_{\pi_\theta,Q}$ the probability law induced by $\pi_\theta$ on $Q$ : $p_{\pi_\theta,Q} = \int_{Q_i} d\pi_\theta$.

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

$$\arg\max \sum_{n=1}^{N} \sum_{L} \hat{p}_{Q^{(n)}} \log B_{l}^{(n)} w$$

$B_{l}^{(n)}$ the $l$-th row of a $(L \times K)$ binary matrix describing the link between observations $Q_i^{(n)}$ and elements of $P$.
- optimization very fast with the modification of Vertex Exchange Method (Hormann et al., Statistics & Probability Letters (2007)).

**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 $Q \rightarrow$ grades
- From observations of grades $G_i$ we only know that $\theta_i \in \{ \theta : Q(B(\theta, P_i)) = G_i \}$.

---

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

$$\dot{\pi}_\theta = \arg\max_{\pi} h_2(\pi) s.t. \quad \left\| (\pi^{(n)} - \hat{p}_{\pi_\theta,Q^{(n)}}) \right\|_{\epsilon}^2 \leq \epsilon^* \forall n,$$

$h_2(\cdot)$ : Rényi-entropy of order 2; $\Sigma^{(n)}$ : $L \times L$ covariance matrix of the empirical distribution of $\hat{p}_{\pi_\theta,Q^{(n)}}$.
- We propose to use likelihood to chose the degree of regularization $\tilde{\pi}_{\theta} = \arg\max_{\pi} \mathcal{L}(\pi)$

---

**Results**
- Unstable solution not physically plausible biological model of a population
- Smoother, more plausible and stable, better for prediction
- Leave-one-out-cross-validation, compare kolmogorov and total variation distances between estimated and observed frequencies.