RECENT ADVANCES ON THE CONSISTENCY OF VARIATIONAL BAYES PROCEDURES

PIERRE ALQUIER
CREST, ENSAE, Université Paris Saclay, 5 avenue Henry le Chatelier,
91120 Palaiseau – France
e-mail: pierre.alquier@ensae.fr

While Bayesian methods are popular in statistics and machine learning, their application to massive datasets is often challenging as MCMC algorithms become prohibitively slow when the sample size and/or the dimension grows. Using a variational approximation allows replacing MCMC by an optimization algorithm that can be much faster. Such methods have been applied in computationally demanding applications such as collaborative filtering, image and video processing, NLP, . . .

However, despite very nice results in practice, the theoretical properties of these approximations were not known until recently. In this talk, I will present and discuss some recent results on the consistency of variational approximations. In [1], we derived consistency results for variational approximations of pseudo-posteriors used in machine learning: \( \exp[-\alpha r_n(\theta)] \pi(d\theta) \) (usually referred to as Gibbs posterior) where \( \alpha > 0 \), \( r_n \) is some empirical risk and \( \pi \) is the prior. In [2], using tools from [3], we derived similar results in a statistical setting, for so-called tempered posteriors – that is, \( L(\theta)^{\alpha} \pi(d\theta) \) where \( 0 < \alpha < 1 \), and \( L(\theta) \) is the likelihood. These tempered posteriors are known to be more robust to model misspecification [4]. The proper Bayesian setting (that is, \( \alpha = 1 \)) was studied by [5, 6]. I will discuss the differences between [2], [5] and [6], the main messages of these papers and the remaining open questions.

References