Mathematics > Statistics Theory

Posterior contraction of the population polytope in finite admixture models

XuanLong Nguyen

(Submitted on 1 Jun 2012)

We study the posterior contraction behavior of the latent population structure that arises in admixture models as the amount of data increases. An admixture model --- alternatively known as a topic model --- specifies k populations, each of which is characterized by a Δ^d-valued vector of frequencies for generating a set of discrete values in ${0,1,..., d}$. The population polytope is defined as the convex hull of the k frequency vectors. Under the admixture specification, each of m individuals generates an i.i.d. frequency vector according to a probability distribution defined on the (unknown) population polytope G_0, and then generates n data points according to the sampled frequency vector. Given a prior distribution over the space of population polytopes, we establish rates at which the posterior distribution contracts to G_0, under the Hausdorff metric and a minimum matching Euclidean metric, as the amount of data $m\times n$ tends to infinity. Rates are obtained for the overfitted setting, i.e., when the number of extreme points of G_0 is bounded above by k, and for the setting in which the number of extreme points of G_0 is known. Minimax lower bounds are also established. Our analysis combines posterior asymptotics techniques for the estimation of mixing measures in hierarchical models with elementary arguments in convex geometry.

Comments:	28 pages
Subjects:	**Statistics Theory (math.ST)**; Learning (cs.LG)
Report number:	Technical report 528, Department of Statistics, University of Michigan
Cite as:	**arXiv:1206.0068v1 [math.ST]**

Submission history

From: XuanLong Nguyen [view email]
[v1] Fri, 1 Jun 2012 02:26:58 GMT (34kb)

Which authors of this paper are endorsers?