Computer Science > Data Structures and Algorithms

Learning k-Modal Distributions via Testing

Constantinos Daskalakis, Ilias Diakonikolas, Rocco A. Servedio

(Submitted on 13 Jul 2011 (v1), last revised 15 Jul 2011 (this version, v2))

A k-modal probability distribution over the domain $\{1,...,n\}$ is one whose histogram has at most k "peaks" and "valleys." Such distributions are natural generalizations of monotone ($k=0$) and unimodal ($k=1$) probability distributions, which have been intensively studied in probability theory and statistics.
In this paper we consider the problem of \emph{learning} an unknown k-modal distribution. The learning algorithm is given access to independent samples drawn from the k-modal distribution p, and must output a hypothesis distribution $\hat{p}$ such that with high probability the total variation distance between p and $\hat{p}$ is at most $\eps.$
We give an efficient algorithm for this problem that runs in time $\poly(k,\log (n),1/\eps)$. For $k \leq \tilde{O}(\sqrt{\log n})$, the number of samples used by our algorithm is very close (within an $\tilde{O}(\log(1/\eps))$ factor) to being information-theoretically optimal. Prior to this work computationally efficient algorithms were known only for the cases $k=0,1.$
A novel feature of our approach is that our learning algorithm crucially uses a new \emph{property testing} algorithm as a key subroutine. The learning algorithm uses the property tester to efficiently decompose the k-modal distribution into k (near)-monotone distributions, which are easier to learn.

Subjects: **Data Structures and Algorithms (cs.DS)**; Learning (cs.LG); Statistics Theory (math.ST)

Cite as: **arXiv:1107.2700 [cs.DS]**

(or **arXiv:1107.2700v2 [cs.DS]** for this version)

Submission history

From: Ilias Diakonikolas [view email]

[v1] Wed, 13 Jul 2011 23:26:53 GMT (33kb)

[v2] Fri, 15 Jul 2011 06:05:19 GMT (33kb)