Computer Science > Data Structures and Algorithms

Learning Poisson Binomial Distributions

Constantinos Daskalakis, Ilias Diakonikolas, Rocco A. Servedio

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

We consider a basic problem in unsupervised learning: learning an unknown \emph{Poisson Binomial Distribution} over $\{0,1,...,n\}$. A Poisson Binomial Distribution (PBD) is a sum $X = X_1 + ... + X_n$ of n independent Bernoulli random variables which may have arbitrary expectations. We work in a framework where the learner is given access to independent draws from the distribution and must (with high probability) output a hypothesis distribution which has total variation distance at most $\eps$ from the unknown target PBD.
As our main result we give a highly efficient algorithm which learns to $\eps$-accuracy using $\tilde{O}(1/\eps^3)$ samples independent of n. The running time of the algorithm is \emph{quasilinear} in the size of its input data, i.e. $\tilde{O}(\log(n)/\eps^3)$ bit-operations (observe that each draw from the distribution is a $\log(n)$-bit string). This is nearly optimal since any algorithm must use $\Omega(1/\eps^2)$ samples.
We also give positive and negative results for some extensions of this learning problem.

Subjects:	**Data Structures and Algorithms (cs.DS)**; Learning (cs.LG); Statistics Theory (math.ST)
Cite as:	**arXiv:1107.2702v2 [cs.DS]**

Submission history

From: Ilias Diakonikolas [view email]
[v1] Wed, 13 Jul 2011 23:30:39 GMT (43kb)
[v2] Fri, 15 Jul 2011 06:03:55 GMT (43kb)

Download:

- PDF
- PostScript
- Other formats

Current browse context:
cs.DS
< prev | next >
new | recent | 1107

Change to browse by:
cs
cs.LG
math
math.ST
stat

References & Citations
- NASA ADS

DBLP - CS Bibliography
listing | bibtex
Constantinos Daskalakis
Ilias Diakonikolas
Rocco A. Servedio

Bookmark(what is this?)