Estimating stationary mass, frequency by frequency

Milind Nakul   MNAKUL3@GATECH.EDU
School of Industrial and Systems Engineering
Georgia Institute of Technology
Atlanta, USA

Vidya Muthukumar   VMUTHUKUMAR8@GATECH.EDU
Schools of Electrical and Computer Engineering
and Industrial and Systems Engineering
Georgia Institute of Technology
Atlanta, USA

Ashwin Pananjady   ASHWINPM@GATECH.EDU
Schools of Industrial and Systems Engineering
and Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, USA

Editors: Nika Haghtalab and Ankur Moitra

Abstract

Suppose we observe a trajectory of length n from an exponentially α-mixing stochastic process over a finite but potentially large state space. We consider the problem of estimating the probability mass placed by the stationary distribution of any such process on elements that occur with a certain frequency in the observed sequence. We estimate this vector of probabilities in total variation distance, showing universal consistency in n and recovering known results for i.i.d. sequences as special cases. Our proposed methodology—implementable in linear time—carefully combines the plug-in (or empirical) estimator with a recently-proposed modification of the Good–Turing estimator called WINGIT, which was originally developed for Markovian sequences. En route to controlling the error of our estimator, we develop new performance bounds on WINGIT and the plug-in estimator for exponentially α-mixing stochastic processes. Importantly, the extensively used method of Poissonization can no longer be applied in our non i.i.d. setting, and so we develop complementary tools—including concentration inequalities for a natural self-normalized statistic of mixing sequences—that may prove independently useful in the design and analysis of estimators for related problems. Simulation studies corroborate our theoretical findings.[1]

Acknowledgments

References

Milind Nakul, Vidya Muthukumar, and Ashwin Pananjady. Estimating stationary mass, frequency by frequency. *arXiv preprint arXiv:2503.12808*, 2025.

---

1. Extended abstract. Full version appears as Nakul et al. (2025).