

## Title of abstract

HaiYing Wang <sup>1</sup>

The Optimal Subsampling Method under the A-optimality Criterion (OSMAC) proposed in [1] samples more informative data points with higher probabilities. However, the original OSMAC estimator use inverse of optimal subsampling probabilities as weights in the likelihood function. This reduces contributions of more informative data points and the resultant estimator may lose efficiency. In this paper [2], we propose a more efficient estimator based on OSMAC subsample without weighting the likelihood function. Both asymptotic results and numerical results show that the new estimator is more efficient. In addition, our focus in this paper is inference for the true parameter, while [1] focuses on approximating the full data estimator. We also develop a new algorithm based on Poisson sampling, which does not require to approximate the optimal subsampling probabilities all at once. This is computationally advantageous when available random-access memory is not enough to hold the full data. Interestingly, asymptotic distributions also show that Poisson sampling produces more efficient estimator if the sampling rate, the ratio of the subsample size to the full data sample size, does not converge to zero. We also obtain the unconditional asymptotic distribution for the estimator based on Poisson sampling.

## References

- [1] Haiying Wang, Rong Zhu, and Ma Ping. Optimal subsampling for large sample logistic regression. *Journal of the American Statistical Association*, DOI:10.1080/01621459.2017.1292914, 2017.
- [2] HaiYing Wang. More efficient estimation for logistic regression with optimal subsample. *arXiv preprint arXiv:1802.02698*, 2018.

---

<sup>1</sup>Department of Statistics, University of Connecticut, Storrs, CT 06269, USA, E-mail: [haiying.wang@uconn.edu](mailto:haiying.wang@uconn.edu)