
STATISTICS SEMINAR

UW-Department of Statistics

www.stat.wisc.edu



TITLE: A Modern
Maximum-Likelihood
Approach for High-
Dimensional
Logistic Regression

Speaker:
Pragya Sur

Department of Statistics
Stanford University

Time & Place:

Monday, February

18, 2019 4pm,

Room 133 SMI

Cookies & Coffee @

3:30, Rm 1210 MSC

Abstract: Logistic regression is arguably the most widely used and studied non-linear model in statistics. Classical maximum-likelihood theory based statistical inference is ubiquitous in this context. This theory hinges on well-known fundamental results: (1) the maximum-likelihood-estimate (MLE) is asymptotically unbiased and normally distributed, (2) its variability can be quantified via the inverse Fisher information, and (3) the likelihood-ratio-test (LRT) is asymptotically a Chi-Squared. In this talk, I will show that in the common modern setting where the number of features and the sample size are both large and comparable, classical results are far from accurate. In fact, (1) the MLE is biased, (2) its variability is far greater than classical results, and (3) the LRT is not distributed as a Chi-Square. Consequently, p-values obtained based on classical theory are completely invalid in high dimensions. In turn, I will propose a new theory that characterizes the asymptotic behavior of both the MLE and the LRT under some assumptions on the covariate distribution, in a high-dimensional setting. Empirical evidence demonstrates that this asymptotic theory provides accurate inference in finite samples. Practical implementation of these results necessitates the estimation of a single scalar, the overall signal strength, and I will propose a procedure for estimating this parameter precisely.

