

DEPARTMENT OF STATISTICS



Yu Bai
Senior Research Scientist
Salesforce Research

*Understanding the Under-Coverage Bias in Uncertainty
Estimation and Calibration*

Wednesday, October 13, 2021
11:45 AM EST

Zoom Meeting: Meeting ID: 924 8283 7055
Password: 378160

<https://rutgers.zoom.us/j/92482837055?pwd=TktKVVU2ZlJzZnZQbm1TNUs3K0pmZz09>

Virtual Coffee session before the seminar at 11:30 AM EST

Abstract: Estimating the data uncertainty in regression tasks is often done by learning a quantile function or a prediction interval of the true label conditioned on the input. It is frequently observed that quantile regression---a vanilla algorithm for learning quantiles with asymptotic guarantees---tends to **under-cover** than the desired coverage level in reality. While various fixes have been proposed, a more fundamental understanding of why this under-coverage bias happens in the first place remains elusive.

In this talk, we present a rigorous theoretical study on the coverage of uncertainty estimation algorithms in learning quantiles. We prove that quantile regression suffers from an inherent under-coverage bias, in a vanilla setting where we learn a realizable linear quantile function and there is more data than parameters. More quantitatively, for $\alpha > 0.5$ and small d/n , the α -quantile learned by quantile regression roughly achieves coverage $\alpha - (\alpha - 1/2) \cdot d/n$ regardless of the noise distribution, where d is the input dimension and n is the number of training data. Our theory reveals that this under-coverage bias stems from a certain parameter estimation error in the high-dimensional proportional limit that is not implied by existing theories on quantile regression. We will also talk about a related phenomenon in the problem of classification calibration---the over-confidence of well-specified logistic regression for binary classification, which can be shown using similar techniques.

Bio: Yu Bai is currently a Senior Research Scientist at Salesforce Research. Prior to joining Salesforce, Yu completed his PhD in Statistics at Stanford University. Yu's research interest lies broadly in machine learning and statistics, with recent focus on the theoretical foundations of deep learning, reinforcement learning, representation learning, and uncertainty quantification.

