

Division of Biostatistics Seminar Series

Veracity in Big Data Analysis: A Statistical Approach to Analyzing Crowdsourced Mobile Data

Soumendra Lahiri, PhD

Stanley A. Sawyer
Professor in Mathematics
and Statistics
Washington University in
St. Louis



When: Friday, April 30, 2021
12:30pm - 1:30 pm

**Registration Link
via Zoom:**

https://wustl.zoom.us/join/registration/tJ0qfu6rqz0rGt1d5JooYMDvkZLJrJEB_1Bc



Veracity in Big Data analysis: A Statistical approach to analyzing Crowdsourced Mobile Data*

S.N. Lahiri

**Department of Mathematics and Statistics
Washington University in St. Louis**

'Big data' is ubiquitous in many areas of applications. While abundance of data provides opportunities for better inference, it also comes with unique challenges, including the issues of volume and veracity of the data that make the task of information extraction very difficult. In this talk, we will consider a crowd-sourced data set collected through an App on mobile devices that gives low quality, high volume, unstructured, noisy measurements on several variables including the ambient temperature. We introduce a way to associate veracity scores to a data value that says how reliable and useful the data value is. We then use this measure of veracity to develop statistical methods for the low quality data and show that it provides significant improvements over naïve applications of the standard prediction methodology. We also explore ways of augmenting information available sources using the crowd-sourced data.

***Joint work with Arnab Chakraborty, Alyson Wilson**