

# STOCHASTIC MODELING IN GENE EXPRESSION MEASUREMENTS: ACCOUNTING FOR BOTH BACKGROUND ERROR AND INSTRUMENT SATURATION

Z.Q. John Lu, National Institute of Standards and Technology, john.lu@nist.gov

There are multiple ways of measuring expressions of thousands of genes simultaneously through measuring the amount of mRNAs by building probes using either the full complementary sequences or using sets of probes pairs based on short subsequences (e.g. Affymetrix arrays). Sources of error include both the non-specific hybridization from using short oligonucleotide arrays to scanner performance limitations due to using the two-color fluorescent dye effects, optical noises, and saturation at high intensity readings. NIST chemists have performed a careful study of two brand scanners over a long period, from which we were able to characterize the scanner effects in the measurement process. In this paper, I describe some statistical issues in fitting a general nonlinear measurement error model for microarray gene expression measurements, which can account for both the background error and instrument saturation, while allowing the signal to vary at a large exponential or Langmuir hyperbolic dynamic range.