

Accelerating Multivariate Simulation Using Graphical Processing Units With Applications to RNA-seq Data

A. Grant Schissler

Department of Mathematics and Statistics, University of Nevada, Reno
aschissler@unr.edu

Abstract

This talk will introduce a parallelized algorithm to facilitate massive, multivariate simulations. A growing number of applications involve multivariate data with dependent marginals. For example, gene expression measurements from biological samples do not behave independently. At the same time, these data are often high dimensional (many measurements per sample) and heterogeneous (differing distributions). These characteristics create a computational challenge when conducting Monte Carlo simulations needed to study the empirical operating characteristics of statistical methodology. Often researchers resort to far-reaching simplifying assumptions that greatly diminish the simulation's usefulness and credibility. To overcome this, we introduce a graphical processing unit (GPU) accelerated version of the well-known Normal to Anything (NORTA) algorithm. In the course of algorithm development, we'll study the role that high-dimensional covariance estimation plays in computational efficiency and statistical properties. Moving to purely computational matters, we'll conduct benchmark studies to elucidate the scenarios that GPU-acceleration produces substantive speed-ups. We'll conclude by deploying the GPU-NORTA algorithm to simulate RNA-sequencing data sets in the context of breast cancer research.

References

1. M.A. SUCHARD AND Q. WANG AND C CHAN AND JACOB FRELINGER AND ANDREW CRON AND AND MIKE WEST.. Understanding GPU Programming for Statistical Computation: Studies in Massively Parallel Massive Mixtures.. *Journal of Computational and Graphical Statistics*, 19(2):419–438, jan 2010..
2. M.C. CARIO AND B.L. NELSON. Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix. Technical report, 1997. Available at <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.48.281&rep=rep1&type=pdf>..
3. C. GENEST AND J. MACKAY.. The Joy of Copulas: Bivariate Distributions with Uniform Marginals. *The American Statistician*, 40(4):280–283, nov 1986..
4. J. WON AND J. LIM AND S. KIM AND AND B. RAJARATNAM. Condition-number-regularized covariance estimation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(3):427–450, jun 2013..