MURI research

Gennady Samorodnitsky

October 2015

Summary of work done in the last year with the emphasis on the last 6 months

In the last year I have been working on several topics.

1. Multivariate Central Limit Theorem for the degree counts (with Sid Resnick).
2. New models for multivariate heavy tails (with a student, Julian Sun).
3. Generating mechanisms for multivariate heavy tails (with the Amherst group, and a student, Emily Fisher).
4. Degree growth for fixed nodes in a network (with Sid Resnick).

1. Multivariate Central Limit Theorem for the degree counts

- This work is within the Statistical Methodology and Multivariate Heavy Tails and Social Network Analysis parts of the grant description.
- The main goal: establish practical confidence regions based on degree counts in real networks.

2. New models for multivariate heavy tails

- This work is within the Statistical Methodology and Mathematical Modeling and Implications parts of the grant description.
- The goal: build practical multivariate models with heavy tails that allow greater flexibility than existing models.

3. Generating mechanisms for multivariate heavy tails

- This work is within the Multivariate Heavy Tails and Social Network Analysis part of the grant description.
- The main goal: design relevant methods to generate multivarate heavy tails.

4. Degree growth for fixed nodes in a network

- This work is within the Multivariate Heavy Tails and Social Network Analysis part of the grant description.
- The main goal: understand the dynamics of growth of nodes with very high degrees.

Some details on the topic **New models for multivariate heavy tails**

Quoting the grant description: "This goals of this MURI project are to generate new classes of multivariate heavy tailed models that highlight the implications of dependence and tail weight and which can be calibrated to data"

Features of heavy tails:

- $P(X > x) \gg e^{-\varepsilon x}$ for any $\varepsilon > 0$.
- $P(X > x + c) \sim P(X > x)$ as x becomes large.

- The most general way of thinking about heavy tailed systems: the main extreme impact on the system comes from individual extreme observations
- The technical **one-dimensional** definition:

$$\frac{P(X_1 + X_2 + \ldots + X_n > x)}{P(X > x)} \to n \quad \text{as } n \to \infty.$$

- The term: a subexponential distribution.

Subexponential distributions allow for:

- power tails (regularly varying tails);
- lognormal-type tails;
- tails of the type

$$P(X > x) \sim \exp\{-cx^{\theta}\}, \; 0 < \theta < 1\,.$$

- This flexibility allows for data calibration.

Multivariate extensions:

- are known in the regularly varying case (via vague convergence on certain spaces of measures);
- our approach is based on the idea of risk regions mentioned in the grant description.
- We need a general description of potentially nonlinear risk regions.

A *risk region* is an open set A in $\mathbb{R}^d$, such that

$$A \text{ is open, increasing, } A^c \text{ convex, } \mathbf{0} \notin \overline{A}.$$

The family of all risk regions is denoted $\mathcal{R}$.

The first step is to define multivariate subexponentiality for a fixed risk region.

- Let $\mathbf{X}$ be a nonnegative random vector.
- Let $A \in \mathcal{R}$.
- Define

$$Y_A = \sup\{u > 0 : \ \mathbf{X} \in uA\}\,.$$

- We say that $\mathbf{X} \in \mathscr{S}_A$ if Y_A is subexponential.
- The class of multivariate subexponential distributions is defined by

$$\mathscr{S}_{\mathcal{R}} = \cap_{A \in \mathcal{R}} \mathscr{S}_A\,.$$

We have a number of criteria for checking that a random vector is multivariate subexponential.

We have proved a theorem that describes the effect of multivariate subexponentiality on risk.

The next step is develop statistical techniques for fitting the new models to data.

Plans for the next year

- Complete the project on generating mechanisms for multivariate heavy tails.
- Complete the project on degree growth for fixed nodes in a network.
- Work on the new project on detecting changes in multivariate heavy tails (with Julian Sun).
- Start a new project with the Columbia group on statistical tools in multivarate heavy tails.