Computational and numerical tools for non-Gaussian multivariate distributions

John Nolan

American University
Washington, DC, USA

MURI Workshop
Soldiers Systems Center
Natick, MA
21 November 2016

Outline

1. Introduction
2. mvmesh Package
   - Directional histograms
3. SimplicialCubature Package
4. SphericalCubature Package
5. gensphere Package

There is a need for non-Gaussian models for multivariate data.
Working in dimension $d > 2$ requires new tools.

- grids and meshes on non-rectangular shapes
- numerical integration over surfaces
- simulate from a shape

There is a need for non-Gaussian models for multivariate data.
Working in dimension $d > 2$ requires new tools.

- grids and meshes on non-rectangular shapes
- numerical integration over surfaces
- simulate from a shape

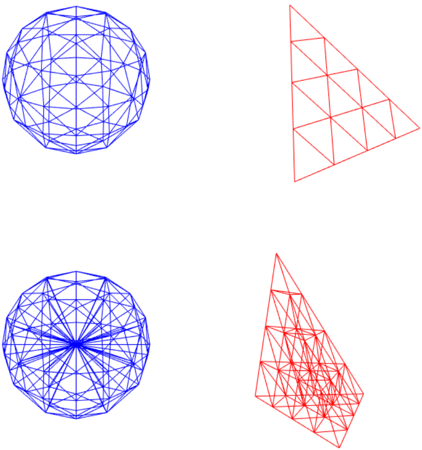
R software packages on open source CRAN

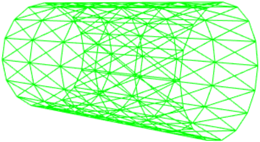
- mvmesh - MultiVariate Meshes
- SphericalCubature
- SimplicialCubature
- gensphere - generalized spherical distributions
- ecdfHT - empirical cdf for Heavy Tailed data
- mvevd - MultiVariate Extreme Value Distributions (in progress)

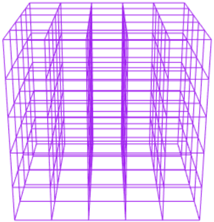
Outline

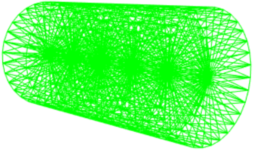
mvmesh

Functions to generate meshes on standard shapes in d dimensions and to work with more complicated shapes

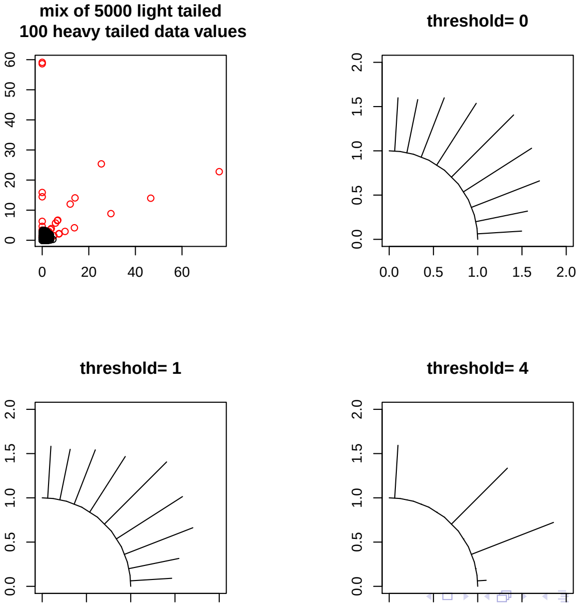






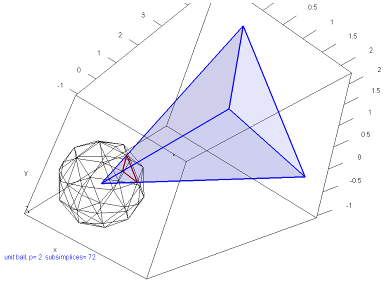


Directional histogram 2D - tabulate # in each cone

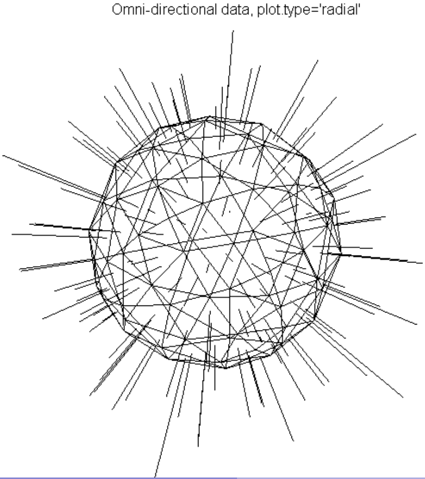


Generalize to $d \geq 3$?

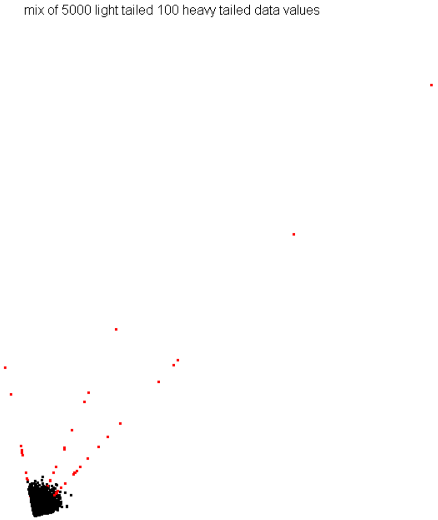
- triangulate sphere
- each simplex on sphere determines a cone
- loop through data points, seeing which cone each falls in
- If $d = 3$, plot
- Variations:
  - threshold based on distance from center
  - use ℓ_p ball
  - restrict to positive orthant



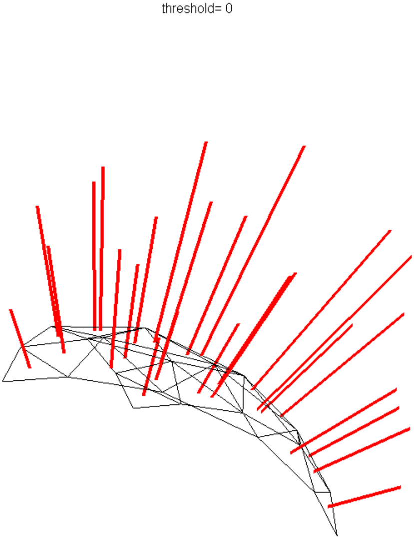
Directional histogram $d = 3$



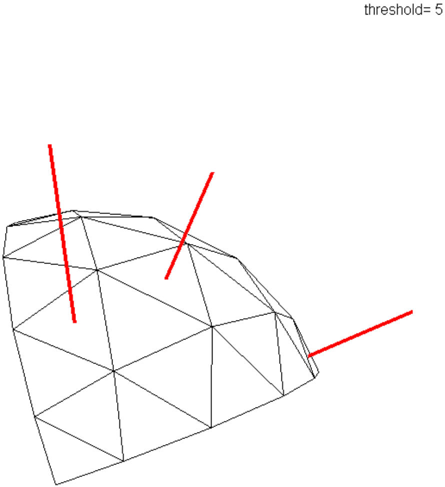
Directional dependence (simulated data)



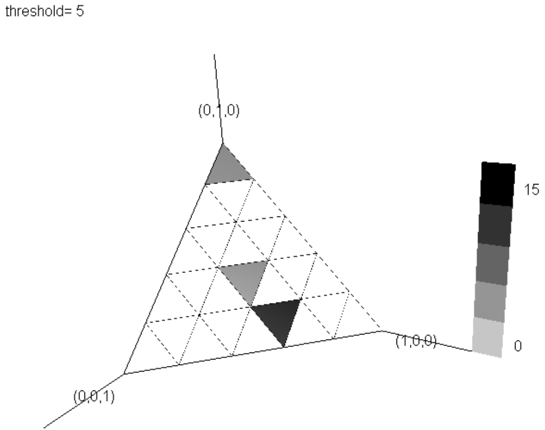
All data



Thresholding by distance from origin



Thresholding by distance from origin (alternate view)



Directional histogram $d > 3$

Subdivision routines return a list of simplices in some order. For any d, can compute the directional histogram counts.

Then plot the a standard histogram using index of simplex.

Directional histogram $d > 3$

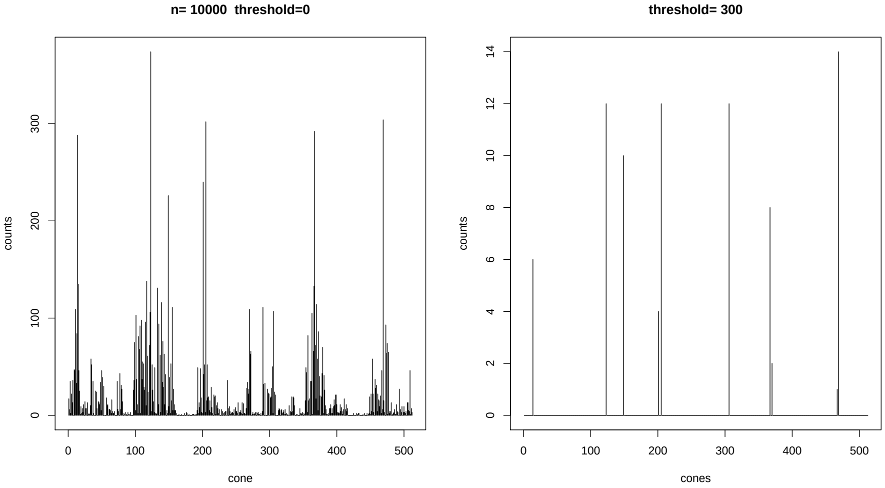
Subdivision routines return a list of simplices in some order. For any d, can compute the directional histogram counts.

Then plot the a standard histogram using index of simplex.

Lose geometry, but can show concentration in different directions. Thresholding may reveal a few directions where extremes lie.

Can use to select model to use on a given data set, e.g. isotropic when histogram is roughly uniform, discrete angular measure when just a few directions present after thresholding.

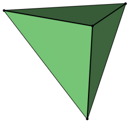
$d = 5$, with 512 cones/directions - $m = 7$ point masses



Outline

Integrating over a simplex

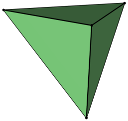
Evaluate $\int_S f(\mathbf{x}) d\mathbf{x}$



where $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ and $S =$ConvexHull$(\mathbf{s}_1, \ldots, \mathbf{s}_{n+1})$ is an n dimensional simplex.

Integrating over a simplex

Evaluate $\int_S f(\mathbf{x}) d\mathbf{x}$



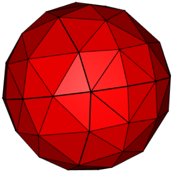
where $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ and $S =$ConvexHull$(\mathbf{s}_1, \ldots, \mathbf{s}_{n+1})$ is an n dimensional simplex.

- exact integration of polynomials using Grundmann-Moler quadrature rules or Lasserre-Avranchenkov algebraic method
- adaptive integration with an R translation of Alan Genz's SimPack, Fortran code. Recursively subdivide simplices.
- extensions to integrate over m-dimensional simplices, $m < n$. Used directly when working with multivariate sum stable, extreme value distributions, and below.

Outline

Integrating over a sphere

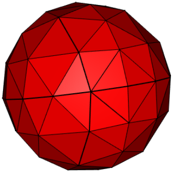
Evaluate $\int_S f(\mathbf{s}) d\mathbf{s}$



where $S = \{\mathbf{s} : |\mathbf{s}| = 1\} \subset \mathbb{R}^n$ is a sphere ($n-1$ dimensional).

Integrating over a sphere

Evaluate $\int_S f(\mathbf{s}) d\mathbf{s}$



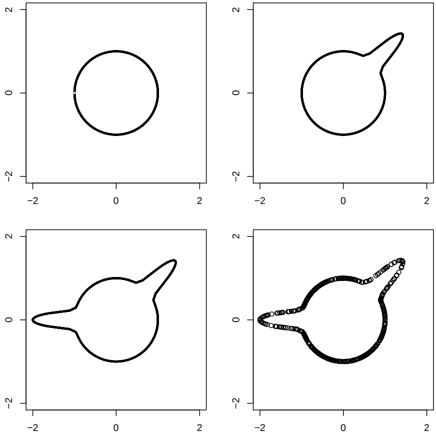
where $S = \{\mathbf{s} : |\mathbf{s}| = 1\} \subset \mathbb{R}^n$ is a sphere ($n - 1$ dimensional).

- exact integration of polynomials
- adaptive integration with using above SimplicialCubature
- extensions to integrate over spherical triangles

Outline

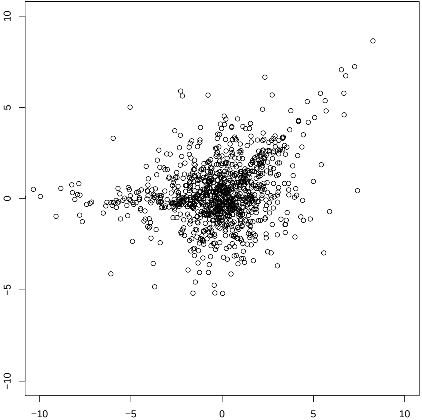
Generalized spherical distributions

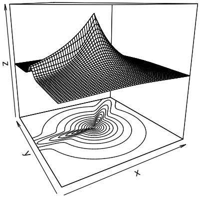
Distributions with level sets that are all scaled versions of a star shaped region. Flexible scheme for building nonstandard star shaped contours.



A tessellation based on the added 'bumps' is automatically generated and used in simulating from the contour. Process requires arclength/surface

Add a radial component to get a distribution: $\mathbf{X} = R\mathbf{Z}$, where $\mathbf{Z}$ is uniform w.r.t. $(d-1)$-dimensional surface area on contour. Here $R \sim \Gamma(2,1)$

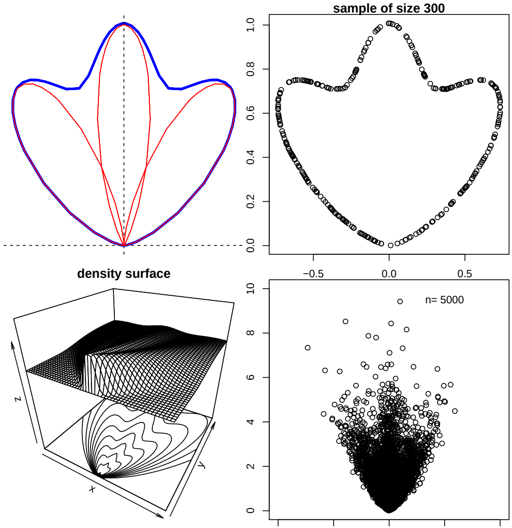




Sample of $\mathbf{X} = R\mathbf{Z}$

density surface

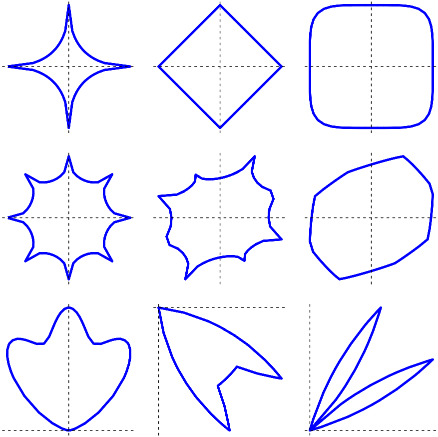
2D example on a cone



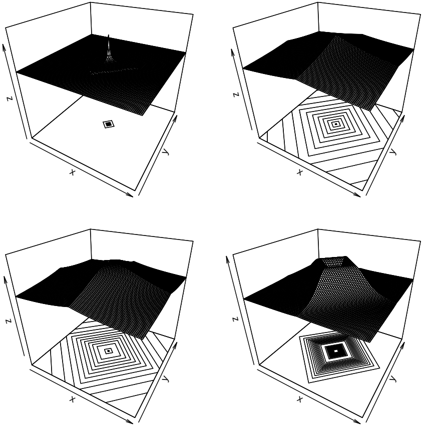
3 Gaussian bumps

Radial $R \sim \Gamma(2, 1)$

Many contour shapes possible



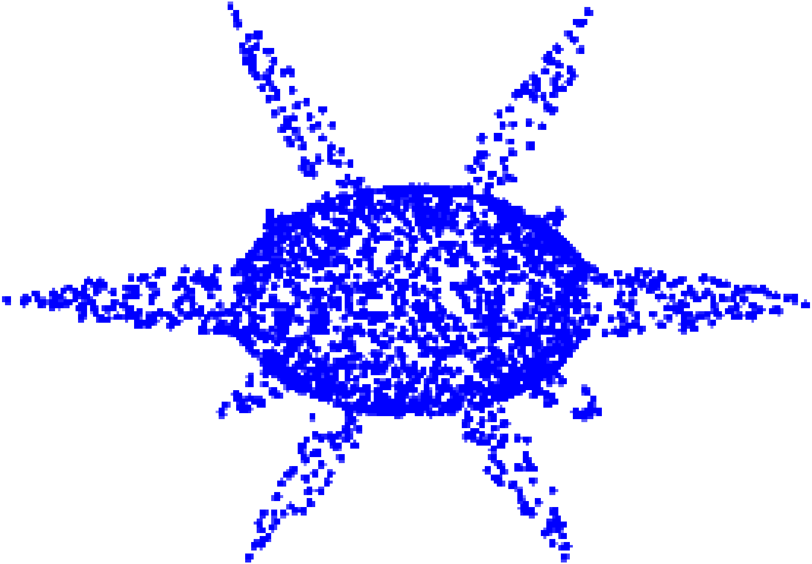
Choice of R determines radial behavior



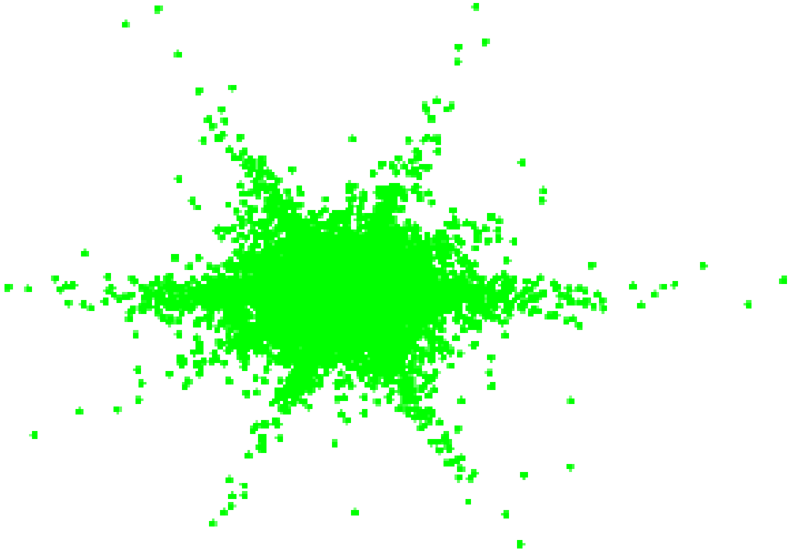
(a) $R \sim$Uniform(0,1)   (b) $R \sim \Gamma(2,1)$ (c) $R = |\mathbf{Y}|$ where $\mathbf{Y}$ is 2D isotropic stable   (d) $R \sim \Gamma(5,1)$

3D example - contour

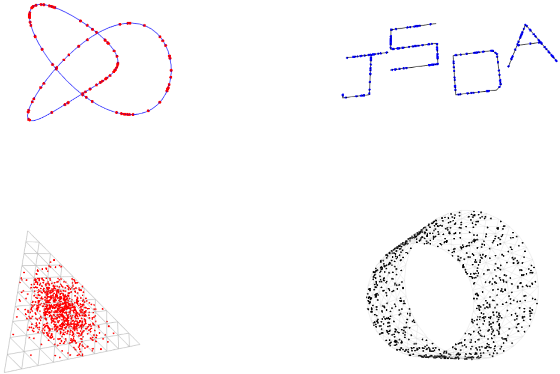
uniform sample from contour



sample from distribution **X** with $R \sim \Gamma(2,1)$



Simulation from general tessellations



Related work

- ecdfHT - empirical cdf for Heavy Tailed data, graphical diagnostic
- flexible classes of multivariate extreme value distributions, partition the unit simplex and put mass in different regions
- flexible classes of multivariate sum stable distributions - partition the unit sphere and put mass in different regions
- refinements of multivariate grids - focus integration routines on specific regions. E.g. compute $P(\mathbf{X} \in S)$ for $\mathbf{X} \sim$ Dirichlet$(\alpha_1, \ldots, \alpha_d)$ and simplex S in the unit simplex.