Software tool for multivariate distributions

John Nolan

American University
Washington, DC, USA

MURI Workshop
Columbia University
New York, NY
16 October 2015

Outline

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

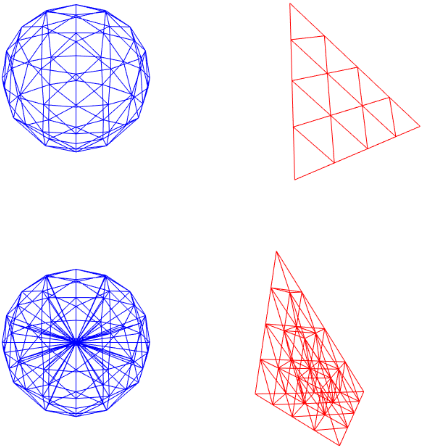
- grids and meshes on non-rectangular shapes
- numerical integration over surfaces
- simulate from a shape

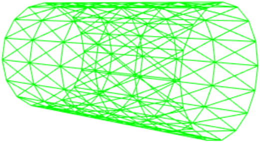
R software packages

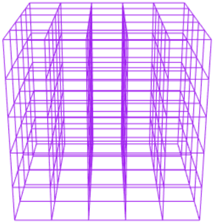
- mvmesh - MultiVariate Meshes (CRAN)
- SphericalCubature (CRAN)
- Simplicial Cubature (CRAN)
- gensphere (manuscript submitted)
- mvevd - MultiVariate Extreme Value Distributions (in progress)

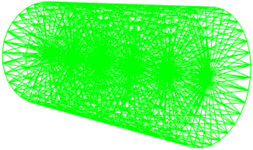
mvmesh

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





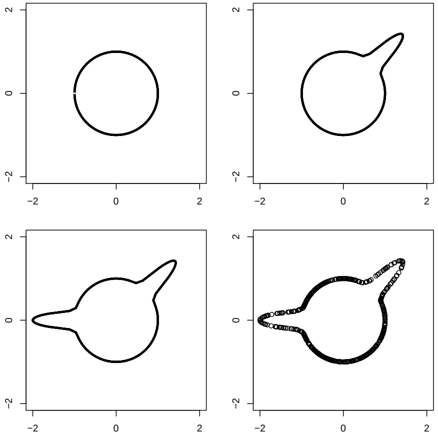




Outline

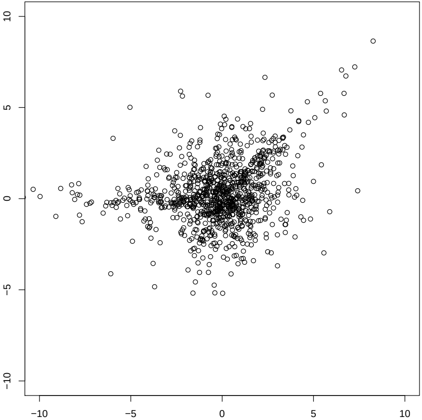
Distributions with level sets that are all scaled versions of a star shaped region.

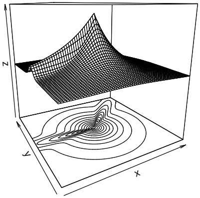
Flexible scheme for building up nonstandard star shaped contours



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

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

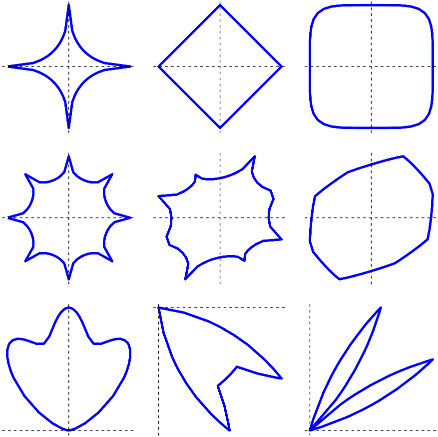




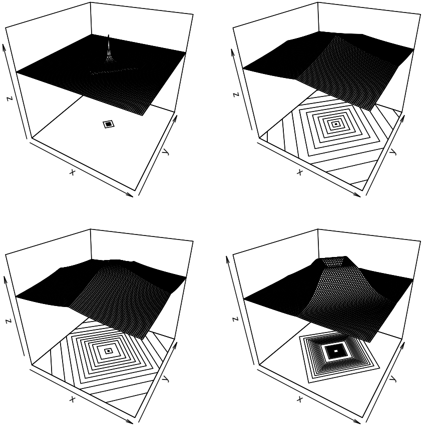
Sample of $\mathbf{X} = R\mathbf{Z}$

density surface

Many contour shapes possible



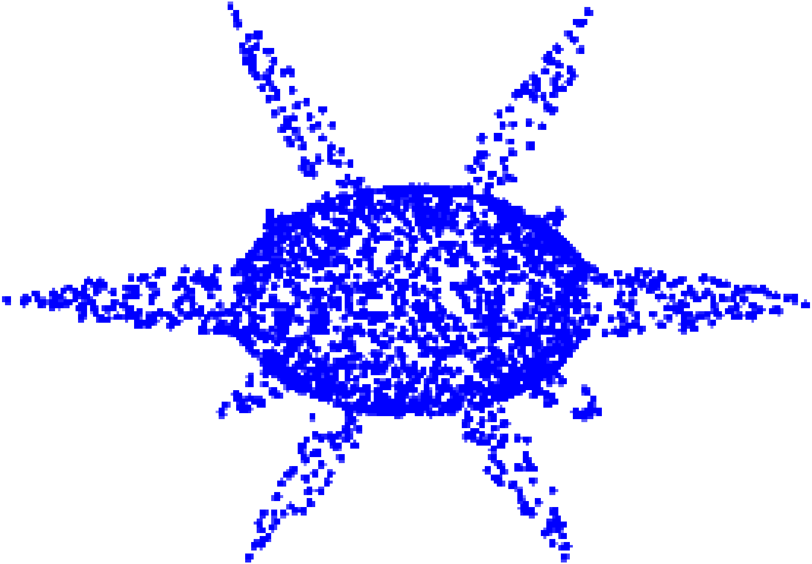
Choice of R determines radial behavior



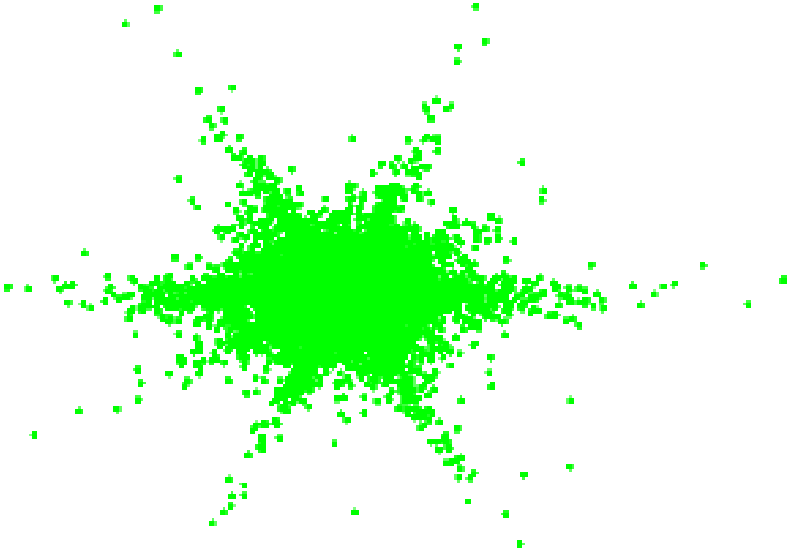
(a) $R \sim$Uniform(0,1) (b) $R \sim \Gamma(2,1)$ (c) $R = |\mathbf{Y}|$ where $\mathbf{Y}$ is 2D isotropic stable (d) $R \sim \Gamma(5,1)$

3D example - contour

uniform sample from contour



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



Outline

Multivariate Fréchet Distributions

de Haan and Resnick (1977): $\mathbf{X}$ max stable, centered with shape index ξ, is characterized by the angular measure H on the unit simplex $\mathbb{W}_+$. The spread of mass by H determines the joint structure. Define the scale function

$$\sigma^{\xi}(\mathbf{u}) = \int_{\mathbb{W}_+} \left(\bigvee_{i=1}^{d} u^{\xi} w_i \right) H(d\mathbf{w}).$$

(If the components of $\mathbf{X}$ are normalized and $\xi = 1$, then this is the tail dependence function $\ell(\mathbf{u})$.)

Multivariate Fréchet Distributions

de Haan and Resnick (1977): $\mathbf{X}$ max stable, centered with shape index ξ, is characterized by the angular measure H on the unit simplex $\mathbb{W}_+$. The spread of mass by H determines the joint structure. Define the scale function

$$\sigma^\xi(\mathbf{u}) = \int_{\mathbb{W}_+} \left(\bigvee_{i=1}^{d} u^\xi w_i \right) H(d\mathbf{w}).$$

(If the components of $\mathbf{X}$ are normalized and $\xi = 1$, then this is the tail dependence function $\ell(\mathbf{u})$.) The scale function determines the joint distribution:

$$G(\mathbf{x}) = P(\mathbf{X} \leq \mathbf{x}) = \exp\left(-\sigma^\xi(\mathbf{x}^{-1}) \right).$$

Observation: need to (a) describe different types of measures and (b) integrate over a surface

R package `mvevd`, $d \geq 2$

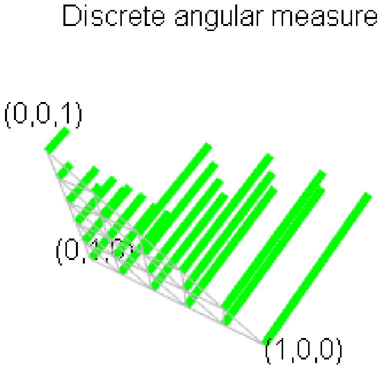
- Define classes of mvevds: discrete H, generalized logistic, Dirichlet mixture, piecewise constant and linear angular measures (computational geometry)
- Compute scale functions $\sigma(\mathbf{u})$ for above classes (integrate over simplices, computational geometry)
- Fitting mvevd data with any of the above classes (max projections)
- Exact simulation from these classes (Dirichlet mix - Dombry, Engelke & Oesting (EVA 2015), Dieker and Mikosch (2015))
- Compute cdf $G(\mathbf{x}) = P(\mathbf{X} \leq \mathbf{x}) = \exp(-\sigma^{\xi}(\mathbf{x}^{-1}))$, $(\boldsymbol{\mu} = 0, \mathbf{x} \geq 0)$.
- Computation of density $g(\mathbf{x})$ when known (partitions)
- Computation of $H(S)$ for a simplex S to estimate tail probabilities in the direction S. (computational geometry & integrate over simplices)

Discrete Angular measures

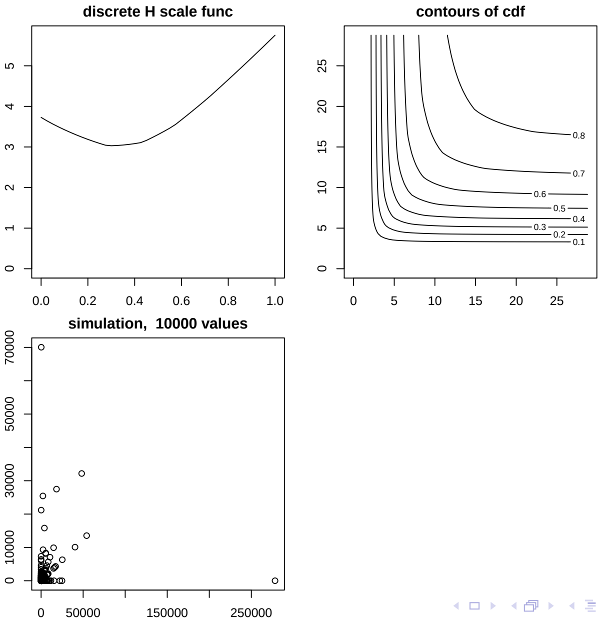
$H(\cdot)$ has mass h_i at points $\mathbf{w}_i$, $i = 1, \ldots, m$;
$\sigma(\mathbf{u}; h_1, \ldots, h_, \mathbf{w}_1, \ldots, \mathbf{w}_m) = \sum_{i=1}^{m} \left(\bigvee_{j=1}^{d} u_j w_{i,j} \right) h_i$.
Sometimes there is a factor model justifying such a choice, or the structure of the problem (Kluppelberg and Gissibl, EVA 2015). In general, this is a dense class, explicit formulas, can simulate exactly.

Discrete Angular measures

$H(\cdot)$ has mass h_i at points $\mathbf{w}_i$, $i = 1, \ldots, m$;
$\sigma(\mathbf{u}; h_1, \ldots, h_{,}\mathbf{w}_1, \ldots, \mathbf{w}_m) = \sum_{i=1}^{m} \left(\bigvee_{j=1}^{d} u_j w_{i,j} \right) h_i.$
Sometimes there is a factor model justifying such a choice, or the structure of the problem (Kluppelberg and Gissibl, EVA 2015). In general, this is a dense class, explicit formulas, can simulate exactly.



2D example with H discrete



Generalized logistic distribution

Fougères, Mercadier, N. (2013) showed that if $\mathbf{S} = (S_1, \ldots, S_d)$ is a positive α-stable random vector with spectral measure Λ concentrated on positive orthant, $0 < \alpha < 1$, and $\mathbf{Z} = (Z_1, \ldots, Z_d)$ is a vector of i.i.d. Frechet($\xi, \mu = 0.\sigma = 1$) components, then

$$\mathbf{X} = \mathbf{S}^{1/\xi}\mathbf{Z} = (S_1^{1/\xi} Z_1, \ldots, S_d^{1/\xi} Z_d)$$

is Frechet($\alpha\xi, \boldsymbol{\mu} = 0, \sigma(\cdot)$ with scale function

$$\sigma^{\alpha\xi}(\mathbf{u}) = c_\alpha \gamma^\alpha(\mathbf{u}^\xi) = c_\alpha \int_{\mathbb{S}} |\langle \mathbf{u}^\xi, \mathbf{s} \rangle|^\alpha \Lambda(d\mathbf{s}).$$

When Λ is discrete and $\alpha\xi = 1$, this simplifies to a sum:

$$\sigma(\mathbf{u}) = c_\alpha \sum_{j=1}^{m} \left| \left\langle \mathbf{u}^\xi, \mathbf{s}_j \right\rangle \right|^\alpha \lambda_j.$$

Generalized logistic cdf and pdf

Formula for the cdf $G(\mathbf{x})$ and density $g(\mathbf{x})$ of $\mathbf{X}$ when Λ is discrete:

$$G(\mathbf{x}) = \exp(-\sigma^{\alpha}(\mathbf{x}^{-\xi})),$$

$$g(\mathbf{x}) = \left\{ \sum_{\pi \in \Pi} (-1)^{|\pi|+d} \prod_{B \in \pi} \frac{\partial^{|B|} I(\mathbf{x})}{\partial^{B} \mathbf{x}} \right\} \times G(\mathbf{x}) \,,$$

$$\frac{\partial^{|B|} I(\mathbf{x})}{\partial^{B} \mathbf{x}} = c_{\alpha} \frac{\alpha!}{(\alpha - |B|)!} \xi^{|B|} \sum_{j=1}^{m} \lambda_j \langle \mathbf{x}^{-\xi}, \mathbf{s}_j \rangle^{\alpha - |B|} \prod_{i \in B} s_{ji} \, x_i^{-\xi - 1},$$

the sum is over Π the set of all partitions of $\{1, \ldots, d\}$ and the product is over all of the blocks B of a partition $\pi \in \Pi$. The number $|\pi|$ denotes the number of blocks of the partition and the cardinality of each block is denoted by $|B|$.

Generalized logistic cdf and pdf

Formula for the cdf $G(\mathbf{x})$ and density $g(\mathbf{x})$ of $\mathbf{X}$ when Λ is discrete:

$$G(\mathbf{x}) = \exp(-\sigma^{\alpha}(\mathbf{x}^{-\xi})),$$

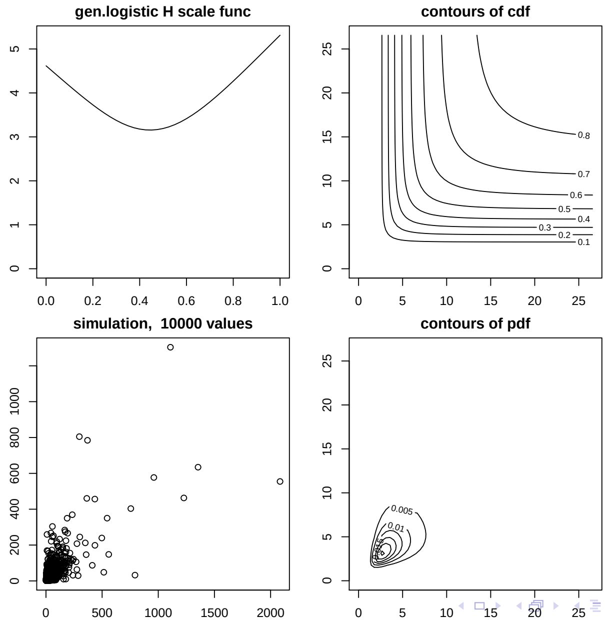
$$g(\mathbf{x}) = \left\{ \sum_{\pi \in \Pi} (-1)^{|\pi|+d} \prod_{B \in \pi} \frac{\partial^{|B|} I(\mathbf{x})}{\partial^{B} \mathbf{x}} \right\} \times G(\mathbf{x}) \ ,$$

$$\frac{\partial^{|B|} I(\mathbf{x})}{\partial^{B} \mathbf{x}} = c_{\alpha} \frac{\alpha!}{(\alpha - |B|)!} \xi^{|B|} \sum_{j=1}^{m} \lambda_j \langle \mathbf{x}^{-\xi}, \mathbf{s}_j \rangle^{\alpha - |B|} \prod_{i \in B} s_{ji} \, x_i^{-\xi-1},$$

the sum is over Π the set of all partitions of $\{1, \ldots, d\}$ and the product is over all of the blocks B of a partition $\pi \in \Pi$. The number $|\pi|$ denotes the number of blocks of the partition and the cardinality of each block is denoted by $|B|$.

The evaluation of this requires the ability to work with all partitions, e.g. R package `partitions`. Code works in any dimension d. In principle can do numerical maximum likelihood estimation, though slow.

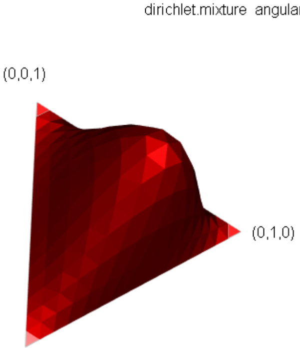
2D example generalized logisitic



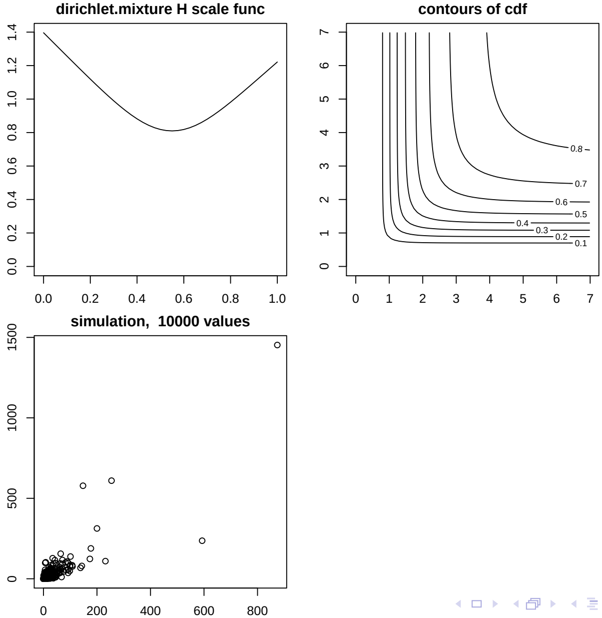
Dirichlet mixture angular measures

Coles and Tawn (1991), Boldi and Davison (2007), Sabourin and Naveau (2014)

$$H(d\mathbf{w}) = \left(\sum_j c_j f(\mathbf{w}; \alpha_j) \right) d\mathbf{w}, \quad f(\cdot; \alpha) \quad \text{Dirichlet density}$$



2D example Dirichlet mix



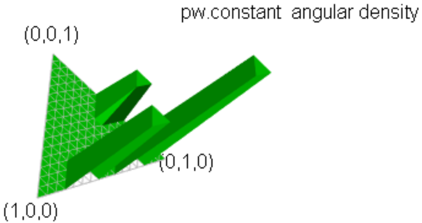
Piecewise polynomial angular densities

Here we need to evaluate integrals of the form

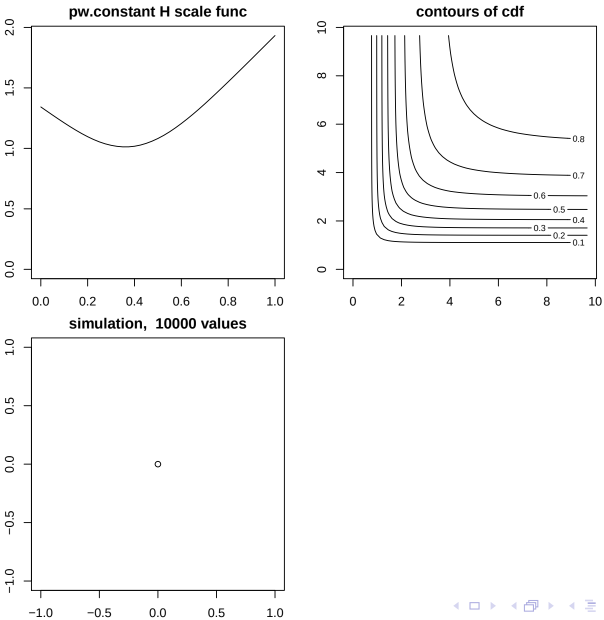
$$\int_{\mathbb{W}_+} \left(\bigvee_{j=1}^{d} u_j w_j \right) h(\mathbf{w}) d\mathbf{w},$$

where $H(d\mathbf{w}) = h(\mathbf{w})d\mathbf{w}$ with angular density $h(\mathbf{w})$ given by a piecewise polynomial.
Can exactly integrate this with careful decomposition of the unit simplex.

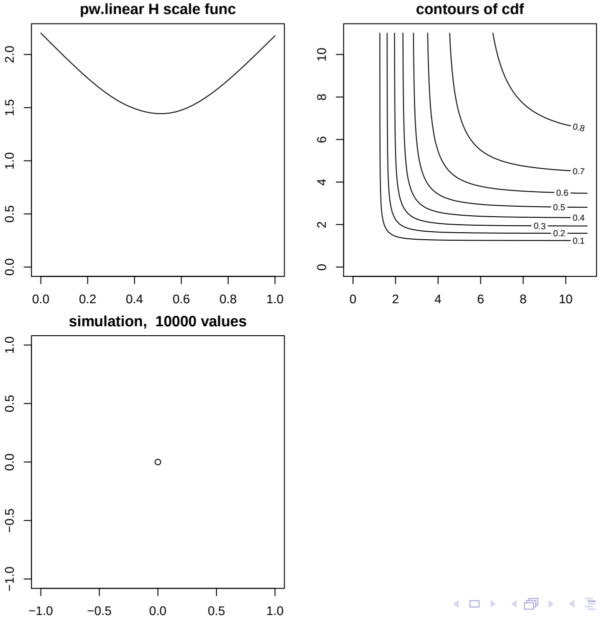
PW constant h



2D example H piecewise constant



2D example H piecewise linear



PW linear h