Directional histograms
Measuring independence for stable distributions

John Nolan
American University
Washington, DC, USA

MURI Workshop, NYC
6 March 2015
1. Directional histograms

2. Independence measure $\eta_p$ for bivariate stable r. vectors

3. Sample measure $\hat{\eta}_p$
Outline

1. Directional histograms

2. Independence measure $\eta_p$ for bivariate stable r. vectors

3. Sample measure $\hat{\eta}_p$
Directional histogram $d = 2$ - count how many in each "direction"

- Mix of 5000 light-tailed data values
- 100 heavy-tailed data values

- Threshold = 0
- Threshold = 1
- Threshold = 4
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 $\ell_p$ ball
  - restrict to positive orthant

Nolan (American U)
MURI NYC
6 March 2015
Directional histogram $d = 3$

Omni-directional data, plot type='radial'
Directional dependence (simulated data)

mix of 5000 light tailed, 100 heavy tailed data values
All data

threshold = 0
Thresholding by distance from origin

threshold = 5
Thresholding by distance from origin (alternate view)

threshold = 5
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 - isotropic

\[\text{counts} \times \text{cone} \]

\[n=10000 \quad \text{threshold}=0\]

\[\text{counts} \times \text{cones} \]

\[\text{threshold}=3\]
\( d = 5 \), with 512 cones/directions - \( m = 7 \) point masses

\[ n = 10000 \quad \text{threshold}=0 \]

\[ \text{threshold}= 300 \]
$d = 5$, with 512 cones/directions - concentration in sectors
Outline

1. Directional histograms

2. Independence measure $\eta_p$ for bivariate stable r. vectors

3. Sample measure $\hat{\eta}_p$
Spectral measure characterization

We will say $X \sim S(\alpha, \Lambda, \delta; j)$, $j = 0, 1$ if its joint characteristic function is given by

$$\phi(u) = E \exp(i\langle u, X \rangle) = \exp \left( - \int_S \omega(\langle u, s \rangle|\alpha; j) \Lambda(ds) + i\langle u, \delta \rangle \right),$$

where

$$\omega(t|\alpha; j) = \begin{cases} 
|t|^\alpha [1 + i \text{sign} (t) \tan \frac{\pi \alpha}{2} (|t|^{1-\alpha} - 1)] & \alpha \neq 1, j = 0 \\
|t|^\alpha [1 - i \text{sign} (t) \tan \frac{\pi \alpha}{2}] & \alpha \neq 1, j = 1 \\
|t| [1 + i \text{sign} (t) \frac{2}{\pi} \log |t|] & \alpha = 1, j = 0, 1.
\end{cases}$$

The 1-parameterization is more commonly used, but discontinuous in $\alpha$. 0-parameterization is a continuous parameterization.
Projection parameterization

Every one dimensional projection $\langle u, X \rangle = u_1 X_1 + u_2 X_2 + \cdots + u_d X_d$ has a univariate stable distribution, with a constant index of stability $\alpha$ and skewness $\beta(u)$, scale $\gamma(u)$ and shift $\delta(u)$ that depend on the direction $u$. We will call the functions $\beta(\cdot)$, $\gamma(\cdot)$ and $\delta(\cdot)$ the projection parameter functions. They determine the joint distribution via the Cramèr-Wold device, so we can parameterize $X$ by these projection parameter functions: $X \sim S(\alpha, \beta(\cdot), \gamma(\cdot), \delta(\cdot), j)$, $j = 0$ or $j = 1$. In what follows, we will always assume that $X$ has normalized components: $\gamma(1, 0) = \gamma(0, 1) = 1$. We will sometimes use polar notation: $\gamma(\theta) := \gamma(\cos \theta, \sin \theta)$. 

Nolan (American U)  
MURI NYC  
6 March 2015  
17 / 37
Projection parameterization

Every one dimensional projection \( \langle u, X \rangle = u_1 X_1 + u_2 X_2 + \cdots + u_d X_d \) has a univariate stable distribution, with a constant index of stability \( \alpha \) and skewness \( \beta(u) \), scale \( \gamma(u) \) and shift \( \delta(u) \) that depend on the direction \( u \).

We will call the functions \( \beta(\cdot) \), \( \gamma(\cdot) \) and \( \delta(\cdot) \) the projection parameter functions. They determine the joint distribution via the Cramér-Wold device, so we can parameterize \( X \) by these projection parameter functions:

\[
X \sim S(\alpha, \beta(\cdot), \gamma(\cdot), \delta(\cdot); j), \ j = 0 \text{ or } j = 1.
\]
Projection parameterization

Every one dimensional projection \( \langle \mathbf{u}, \mathbf{X} \rangle = u_1 X_1 + u_2 X_2 + \cdots + u_d X_d \) has a univariate stable distribution, with a constant index of stability \( \alpha \) and skewness \( \beta(\mathbf{u}) \), scale \( \gamma(\mathbf{u}) \) and shift \( \delta(\mathbf{u}) \) that depend on the direction \( \mathbf{u} \).

We will call the functions \( \beta(\cdot), \gamma(\cdot) \) and \( \delta(\cdot) \) the projection parameter functions. They determine the joint distribution via the Cramér-Wold device, so we can parameterize \( \mathbf{X} \) by these projection parameter functions:

\[
\mathbf{X} \sim \mathbf{S}(\alpha, \beta(\cdot), \gamma(\cdot), \delta(\cdot); j), \quad j = 0 \text{ or } j = 1.
\]

In what follows, we will always assume that \( \mathbf{X} \) has normalized components:

\( \gamma(1, 0) = \gamma(0, 1) = 1. \)

Will sometimes use polar notation: \( \gamma(\theta) := \gamma(\cos \theta, \sin \theta). \)
\( \Lambda(\cdot) \) and \( \gamma(\cdot) \)

\text{independent}

\[ \gamma^\alpha(\theta), \quad \alpha = 1.5 \]
isotropic

\[ \gamma^\alpha(\theta), \quad \alpha = 1.5 \]
pos. linear dep.

\[ \gamma^\alpha(\theta), \quad \alpha = 1.5 \]
pos. associated

\[ \gamma^\alpha(\theta), \quad \alpha = 1.5 \]
Set $\gamma_\perp(u) = (|u_1|^\alpha + |u_2|^\alpha)^{1/\alpha}$ (independence), $p \in [1, \infty]$

$$\eta_p = \eta_p(X_1, X_2) = \|\gamma^\alpha(u_1, u_2) - \gamma_\perp^\alpha(u_1, u_2)\|_{L^p(S, du)}. \quad (1)$$

Here $du$ is (unnormalized) surface area on $S$. 
Set $\gamma_\perp(u) = (|u_1|^\alpha + |u_2|^\alpha)^{1/\alpha}$ (independence), $p \in [1, \infty]$.

$$\eta_p = \eta_p(X_1, X_2) = \|\gamma^\alpha(u_1, u_2) - \gamma_\perp^\alpha(u_1, u_2)\|_{L^p(S, du)}.$$

Here $du$ is (unnormalized) surface area on $S$.

$X$ has independent components if and only if $\eta_p = 0$ for some (every) $p \in [1, \infty]$.

$\eta_p$ measures how far the scale function of $X$ is from the scale function of a stable r. vector with independent components: when $X$ is symmetric, earlier work shows $\sup_{x \in \mathbb{R}^2} |f(x) - f_\perp(x)| \leq k_\alpha \|\gamma(\cdot) - \gamma_\perp(\cdot)\|$. 
\[ |\gamma_1^\alpha(\theta) - \gamma_2^\alpha(\theta)| \]
**Properties of $\eta_p$**

- The $p$-norm in (1) is evaluated as an integral over the unit circle $\mathbb{S}$, not all of $\mathbb{R}^2$. In polar coordinates,

$$
\eta_p = \left(2 \int_{0}^{\pi} |\gamma^\alpha(\cos \theta, \sin \theta) - \gamma^\alpha_\perp(\cos \theta, \sin \theta)|^p \, d\theta \right)^{1/p},
$$

where the interval of integration has been reduced by using the fact that $\gamma(\cdot)$ is $\pi$-periodic.

- $\alpha$ can be any value in $(0, 2)$ and $X$ can have symmetric or non-symmetric components, and it can be centered or shifted.

- $\eta_p$ is symmetric: $\eta_p(X_1, X_2) = \eta_p(X_2, X_1)$. 


• $\eta_p \geq 0$ by definition, not measuring positive/negative dependence, just distance from independence. Don't think there is a general way of assigning a sign, e.g. rotate the indep. components case by $\pi/4$ and the resulting distribution bunches around both the lines $y = x$ and $y = -x$ for large values of $|X|$.

• The definition makes sense in the Gaussian case: when $\alpha = 2$, the scale function for a bivariate Gaussian distribution with correlation $\rho$ is $\gamma(u)^2 = 1 + 2\rho u_1 u_2$ and $\gamma_\perp = 1$. Then $\eta_P^p = |2\rho|^p \int_S |u_1 u_2|^p du$, so $\eta_p = k_p |\rho|$. In elliptically contoured/sub-Gaussian case, can get an integral expression that can be evaluated numerically.

• Multivariate stable $X = (X_1, \ldots, X_d)$ has mutually independent components if and only if all pairs are independent, so the components of $X$ are mutually independent if and only if $\eta_p(X_i, X_j) = 0$ for all $i > j$. 
Covariation and co-difference in terms of $\gamma(\cdot)$

For $\alpha > 1$, the covariation is

$$[X_1, X_2]_\alpha = \int_S s_1 s_2^{\langle \alpha-1 \rangle} \Lambda(ds) = \frac{1}{\alpha} \left. \frac{\partial \gamma^\alpha(u_1, u_2)}{\partial u_1} \right|_{(u_1=0, u_2=1)}.$$ 

Thus the covariation depends only on the behavior of $\gamma(\cdot, \cdot)$ near the point $(1, 0)$. If $X_1$ and $X_2$ are independent, then $[X_1, X_2]_\alpha = 0$; but the converse is false.
Covariation and co-difference in terms of $\gamma(\cdot)$

For $\alpha > 1$, the covariation is

$$[X_1, X_2]_\alpha = \int_{S} s_1 s_2^{<\alpha-1>} \wedge(ds) = \frac{1}{\alpha} \frac{\partial \gamma^\alpha(u_1, u_2)}{\partial u_1} \bigg|_{(u_1=0, u_2=1)}.$$

Thus the covariation depends only on the behavior of $\gamma(\cdot, \cdot)$ near the point $(1, 0)$. If $X_1$ and $X_2$ are independent, then $[X_1, X_2]_\alpha = 0$; but the converse is false.

The co-difference is defined for symmetric $\alpha$-stable vectors, and can be written as

$$\tau = \tau(X_1, X_2) = \gamma^\alpha(1, 0) + \gamma^\alpha(0, 1) - \gamma^\alpha(1, -1),$$

and is defined for any $\alpha \in (0, 2)$. If $X_1$ and $X_2$ are independent, then $\tau = 0$. When $\alpha < 1$ and $\tau = 0$, then indep. If $\alpha > 1$, need both $\tau(X_1, X_2) = 0$ and $\tau(X_2, X_1) = 0$ to guarantee indep.
Outline

1. Directional histograms

2. Independence measure $\eta_p$ for bivariate stable r. vectors

3. Sample measure $\hat{\eta}_p$
Use max. likelihood estimation of the marginals and get $\hat{\alpha}$, normalize each component. For angles $0 \leq \theta_1 < \theta_2 < \cdots < \theta_m \leq \pi$, define

$\hat{\gamma}_j = \hat{\gamma}(\cos \theta_j, \sin \theta_j) = \text{ML estimate of the scale of the projected data set } \langle Y_i, (\cos \theta_j, \sin \theta_j) \rangle, \ i = 1, \ldots, n$
Use max. likelihood estimation of the marginals and get $\hat{\alpha}$, normalize each component. For angles $0 \leq \theta_1 < \theta_2 < \cdots < \theta_m \leq \pi$, define

$$\hat{\gamma}_j = \hat{\gamma}(\cos \theta_j, \sin \theta_j) = \text{ML estimate of the scale of the projected data set } \langle Y_i, (\cos \theta_j, \sin \theta_j) \rangle, \ i = 1, \ldots, n$$

Define

$$\hat{\eta}_2 = \left( \sum_{j=1}^{m} \left( \hat{\gamma}_j \hat{\alpha}_j - \gamma_{\perp j} \right)^2 \right)^{1/2}.$$

Get critical values by simulation, depends on $\alpha$ and grid.

Suggest uniform grid with $m$ points in first and second quadrant that avoid 0, $\pi/2$, $\pi$
Uniform grid with $m = 3$ in each quadrant
Covariance of $\hat{\gamma}(\theta_1)$ and $\hat{\gamma}(\theta_2)$
Power calculation via simulation, $\alpha = 1.5$, 5 grid points per quadrant, 1000 simulations

<table>
<thead>
<tr>
<th>$n$</th>
<th>isotropic</th>
<th>indep. $\pi/4$</th>
<th>indep. $\pi/8$</th>
<th>indep. $\pi/16$</th>
<th>exact linear dep.</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.191</td>
<td>0.322</td>
<td>0.243</td>
<td>0.213</td>
<td>1</td>
</tr>
<tr>
<td>50</td>
<td>0.223</td>
<td>0.624</td>
<td>0.381</td>
<td>0.183</td>
<td>1</td>
</tr>
<tr>
<td>100</td>
<td>0.344</td>
<td>0.918</td>
<td>0.644</td>
<td>0.214</td>
<td>1</td>
</tr>
<tr>
<td>200</td>
<td>0.636</td>
<td>0.998</td>
<td>0.937</td>
<td>0.440</td>
<td>1</td>
</tr>
<tr>
<td>300</td>
<td>0.874</td>
<td>1</td>
<td>0.997</td>
<td>0.627</td>
<td>1</td>
</tr>
<tr>
<td>400</td>
<td>0.960</td>
<td>1</td>
<td>1</td>
<td>0.791</td>
<td>1</td>
</tr>
<tr>
<td>500</td>
<td>0.989</td>
<td>1</td>
<td>1</td>
<td>0.893</td>
<td>1</td>
</tr>
<tr>
<td>600</td>
<td>0.999</td>
<td>1</td>
<td>1</td>
<td>0.959</td>
<td>1</td>
</tr>
<tr>
<td>700</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.980</td>
<td>1</td>
</tr>
<tr>
<td>800</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.985</td>
<td>1</td>
</tr>
<tr>
<td>900</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.998</td>
<td>1</td>
</tr>
<tr>
<td>1000</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.997</td>
<td>1</td>
</tr>
</tbody>
</table>
Multivariate: compute $\hat{\eta}_{i,j}$ between all pairs $(X_i, X_j)$
Time series - plot $\eta(X_t, X_{t+h})$

Simulated data with stable innovations:

AR(1), coef = 0.5
n = 1000   alpha = 1.468   beta = -0.082
Time series - returns of Merck stock for 2010-2014

MRK
n = 1257   alpha = 1.744   beta = -0.078
Robustness of acf vs $\eta$ plot

Simulated time series with independent stable terms. In this simulation, the $\eta$ and acf plots look similar (left). Changing one point by replacing a point 15 time periods away from max with 0.8*max shows $\eta$ plot unchanged, but acf shows strong dependence (right).

Independent
$n=1000$  $\alpha=1.54$  $\beta=0.091$

Independent with one extreme value added
$n=1000$  $\alpha=1.523$  $\beta=0.099$
$\eta$ for $\mathbf{X}$ in the domain of attraction of stable

The calculation of $\eta$ only requires an estimate of the tail index $\alpha$ and scale in directions $\theta_1, \ldots, \theta_m$. Can use any tail estimator of the univariate data sets obtained by projecting the data in different directions. The following examples used a simple tail estimator - regression on the tail probabilities.

Simulated using symmetrized Paretos: $X = Y_1 - Y_2$ where each term is $\text{indep. Pareto}(\alpha = 1.5)$. Fix $n = \text{sample size}$. Find critical value by simulation. Bootstrap indep. components $(X_1, X_2)$, compute $\hat{\eta}$ and tabulate. Repeat $M = 10000$ times and find a critical value $c_p$ based on $(1 - p)$ quantile of tabulated values.

Simulate different data sets: isotropic ($\cos U, \sin U$) $X$ where $U \sim \text{Uniform}(0, 2\pi)$; rotations of independent case $R(\theta)(X_1, X_2)$ for $\theta = \pi/4, \pi/8, \pi/16$; exact linear dependence $\epsilon(X, X)$ where $\epsilon = \pm 1$ w/ prob. 1/2. Vary $n$ and tabulate power.
$\eta$ for $X$ in the domain of attraction of stable

The calculation of $\eta$ only requires an estimate of the tail index $\alpha$ and scale in directions $\theta_1, \ldots, \theta_m$. Can use any tail estimator of the univariate data sets obtained by projecting the data in different directions. The following examples used a simple tail estimator - regression on the tail probabilities.

Simulated using symmetrized Paretos: $X = Y_1 - Y_2$ where each term is indep. Pareto($\alpha = 1.5$).

- Fix $n=$sample size.
- Find critical value by simulation. Bootstrap indep. components $(X_1, X_2)$, compute $\hat{\eta}$ and tabulate. Repeat $M = 10000$ times and find a critical value $c_p$ based on $(1 - p)$ quantile of tabulated values.
- Simulate different data sets: isotropic $(\cos U, \sin U)X$ where $U \sim \text{Uniform}(0, 2\pi)$; rotations of independent case $R(\theta)(X_1, X_2)$ for $\theta = pi/4, pi/8, pi/16$; exact linear dependence $\epsilon(X, X)$ where $\epsilon = \pm 1$ w/ prob. 1/2.
- Vary $n$ and tabulate power
Power calculations in DOA case

<table>
<thead>
<tr>
<th>sample size $n$</th>
<th>isotropic</th>
<th>independent $\sim \pi/4$</th>
<th>independent $\sim \pi/8$</th>
<th>independent $\sim \pi/16$</th>
<th>exact linear dependence</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.253</td>
<td>0.057</td>
<td>0.049</td>
<td>0.058</td>
<td>0.161</td>
</tr>
<tr>
<td>200</td>
<td>0.708</td>
<td>0.025</td>
<td>0.040</td>
<td>0.049</td>
<td>0.342</td>
</tr>
<tr>
<td>300</td>
<td>0.844</td>
<td>0.010</td>
<td>0.013</td>
<td>0.023</td>
<td>0.481</td>
</tr>
<tr>
<td>400</td>
<td>0.940</td>
<td>0.011</td>
<td>0.020</td>
<td>0.022</td>
<td>0.995</td>
</tr>
<tr>
<td>500</td>
<td>0.956</td>
<td>0.011</td>
<td>0.007</td>
<td>0.018</td>
<td>1</td>
</tr>
<tr>
<td>600</td>
<td>0.986</td>
<td>0.024</td>
<td>0.013</td>
<td>0.028</td>
<td>1</td>
</tr>
<tr>
<td>700</td>
<td>0.988</td>
<td>0.023</td>
<td>0.003</td>
<td>0.009</td>
<td>1</td>
</tr>
<tr>
<td>800</td>
<td>0.995</td>
<td>0.258</td>
<td>0.012</td>
<td>0.019</td>
<td>1</td>
</tr>
<tr>
<td>900</td>
<td>0.998</td>
<td>0.284</td>
<td>0.013</td>
<td>0.011</td>
<td>1</td>
</tr>
<tr>
<td>1000</td>
<td>0.993</td>
<td>0.498</td>
<td>0.006</td>
<td>0.009</td>
<td>1</td>
</tr>
<tr>
<td>2000</td>
<td>1</td>
<td>0.996</td>
<td>0.376</td>
<td>0.008</td>
<td>1</td>
</tr>
<tr>
<td>3000</td>
<td>1</td>
<td>1</td>
<td>0.876</td>
<td>0.003</td>
<td>1</td>
</tr>
<tr>
<td>4000</td>
<td>1</td>
<td>1</td>
<td>0.989</td>
<td>0.003</td>
<td>1</td>
</tr>
<tr>
<td>5000</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.004</td>
<td>1</td>
</tr>
</tbody>
</table>

Require larger sample to detect dependence; depends on choosing cutoff correctly and estimators of $\alpha$ and scale.