A new distribution for modeling cylindrical data

University of Shizuoka

Imoto Tomoaki

BOSTON UNIVERSITY / KEIO UNIVERSITYT / SINGHUA UNIVERSITY WORKSHOP 2023

June 30th, 2023

Contents

- Preliminary
  - Directional statistics
  - Circular distribution
  - Cylindrical distribution
- Proposition
  - New distribution family
  - Property
  - Special case
- Illustrative applications
  - Wind data
  - Pass sonar data
- Concluding remarks

Preliminary

Directional statistics

The statistics for dealing with data that has periodicity.

Typical examples of the data are the direction of wind and time in a day.

Fig. The direction of wind at pm 6:00
at Texas (2003/5/20-2003/07/31)

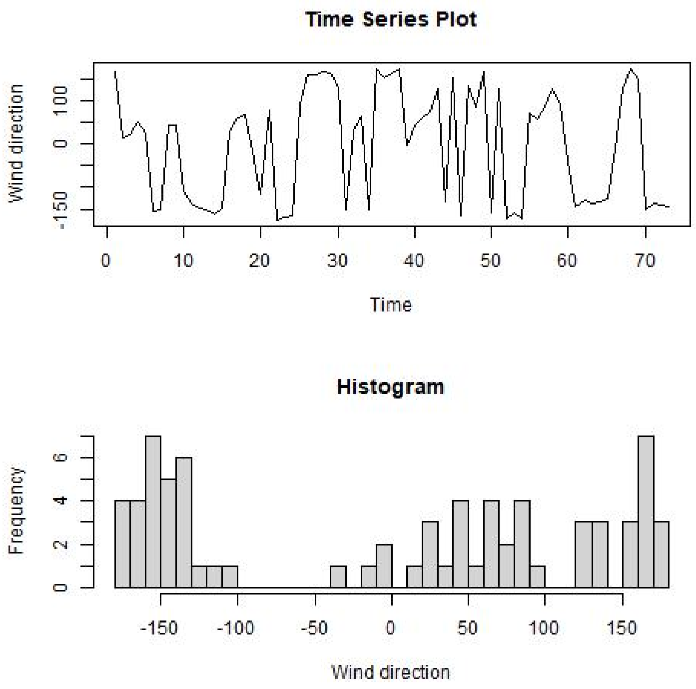


Fig. The time (by o’clock) of thunder heard
at Kew Gardens (1910-1935)

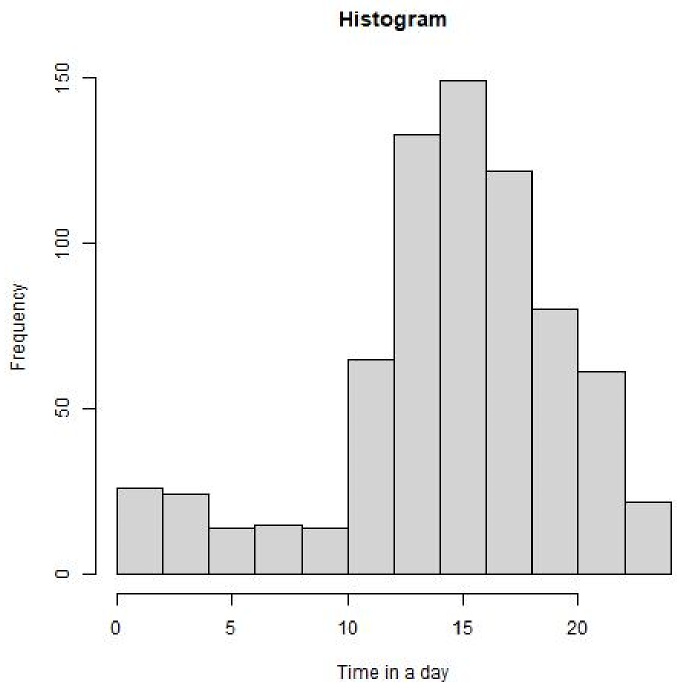


Fig. Direction of wind at pm 6:00
at Texas (2003/5/20-2003/07/31)

(a) 0 = North

(b) 0 = East-North-East

(c) 0 = South-West-South

(c) 0 = South

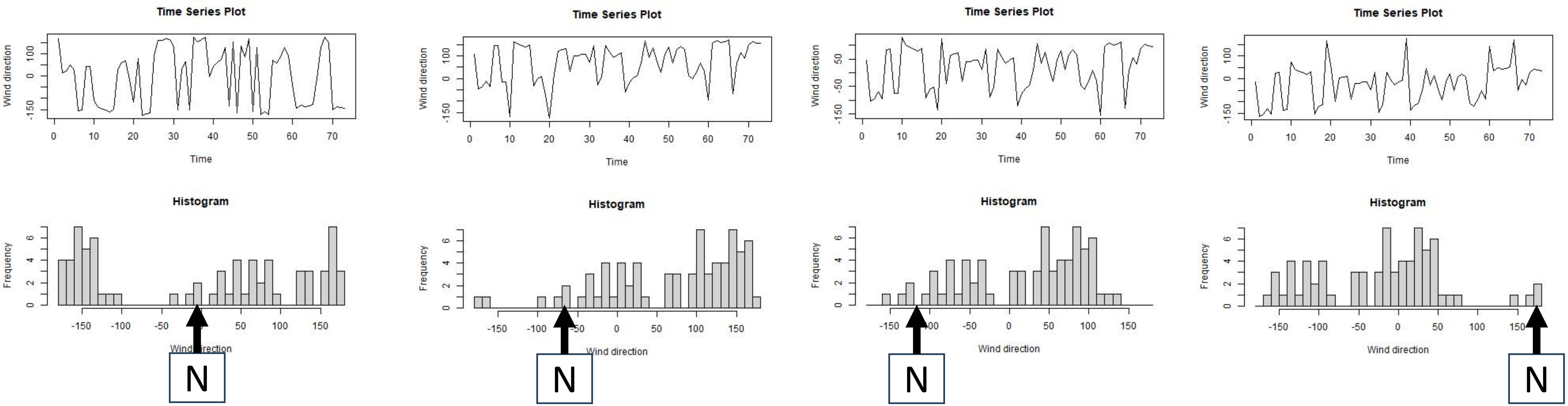
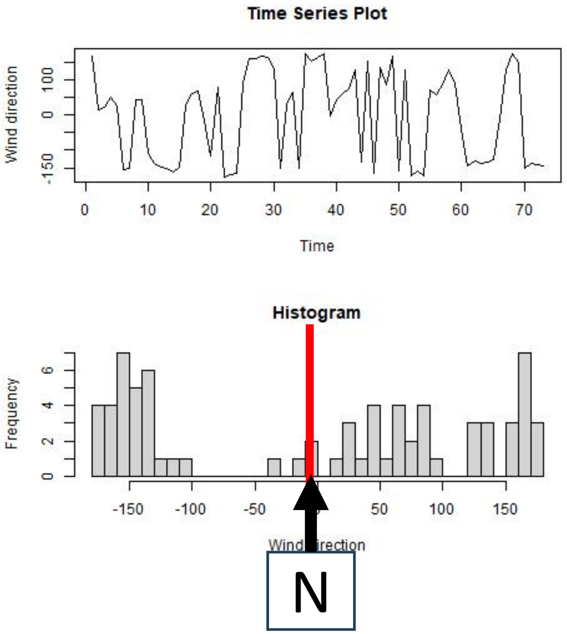


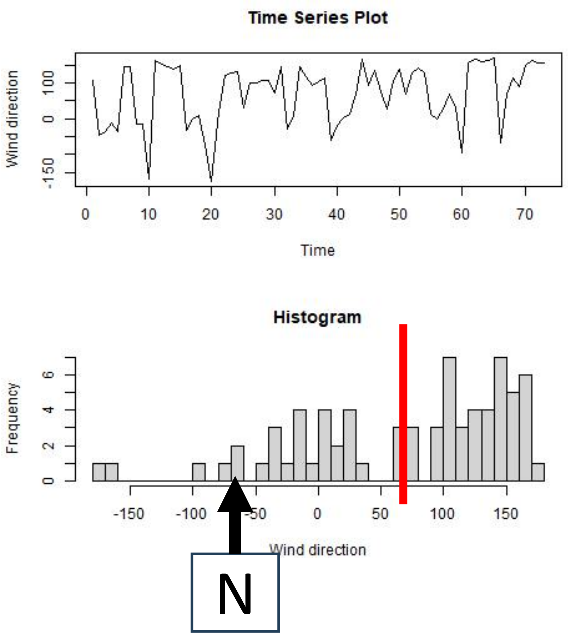
Fig. Direction of wind at pm 6:00
at Texas (2003/5/20-2003/07/31)

(a) 0 = North



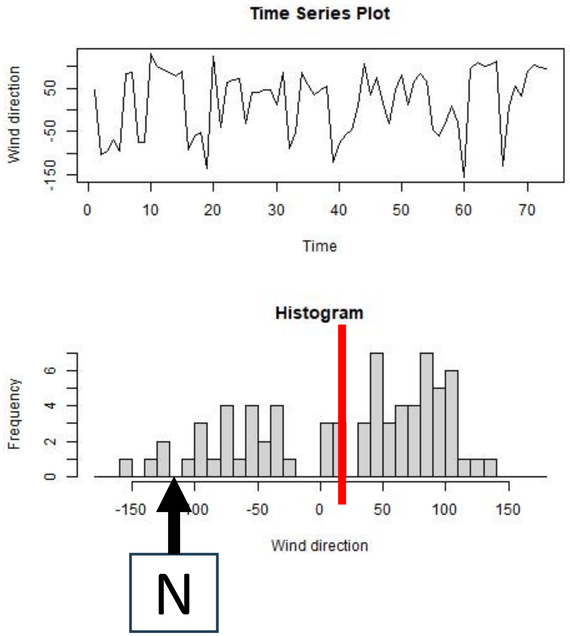
Arithmetic mean:
-3.4° from N

(b) 0 = East-North-East



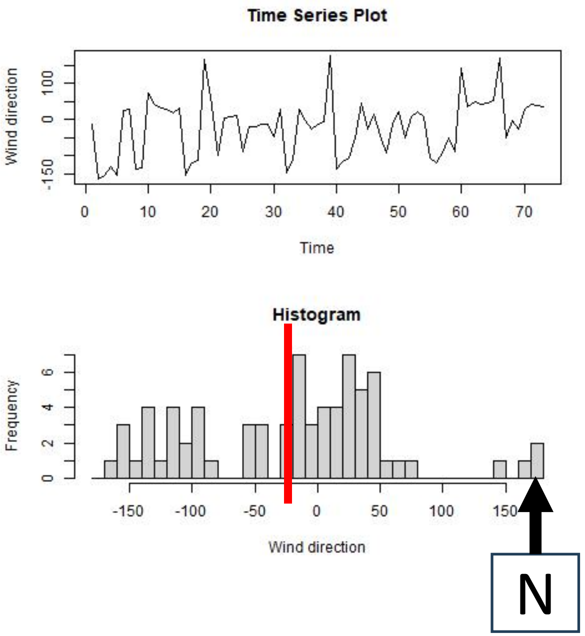
Arithmetic mean:
69.8° from E-N-E

(c) 0 = South-West-South



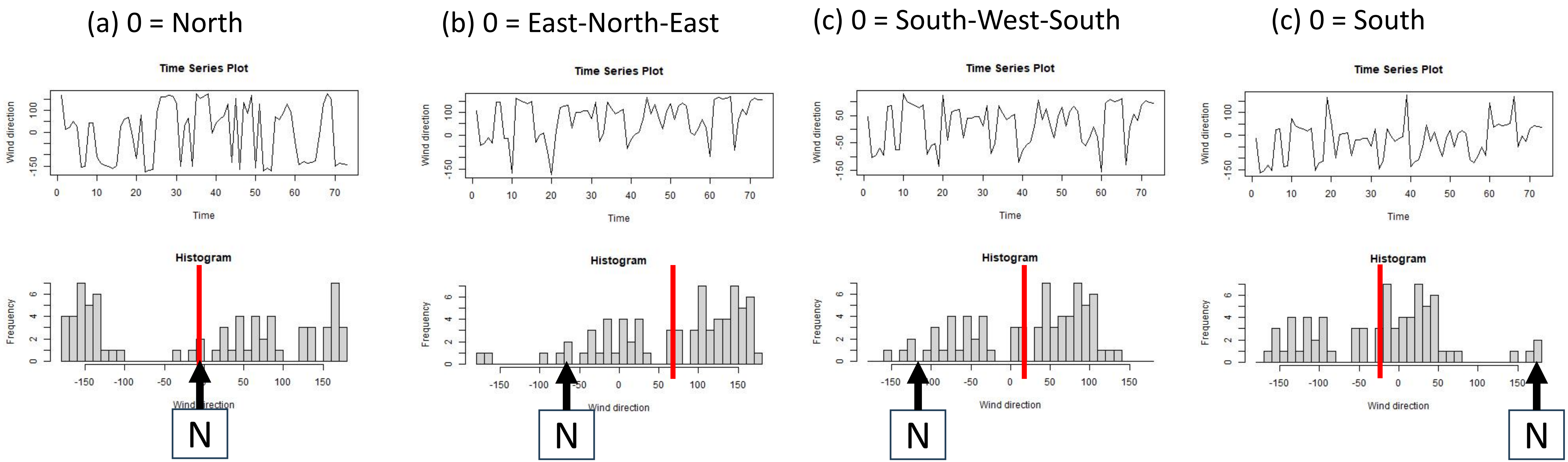
Arithmetic mean:
19.7° from S-W-S

(c) 0 = South



Arithmetic mean:
-20.6° from S

Fig. Direction of wind at pm 6:00
at Texas (2003/5/20-2003/07/31)



In directional statistics, the arithmetic mean cannot become the representative value of data. For example, the observations

$10°$ and $350°$ from North

Mean = $180°$ from North

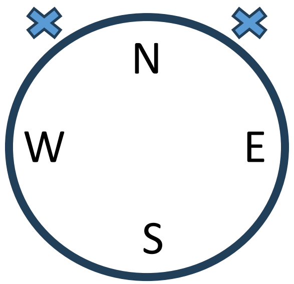


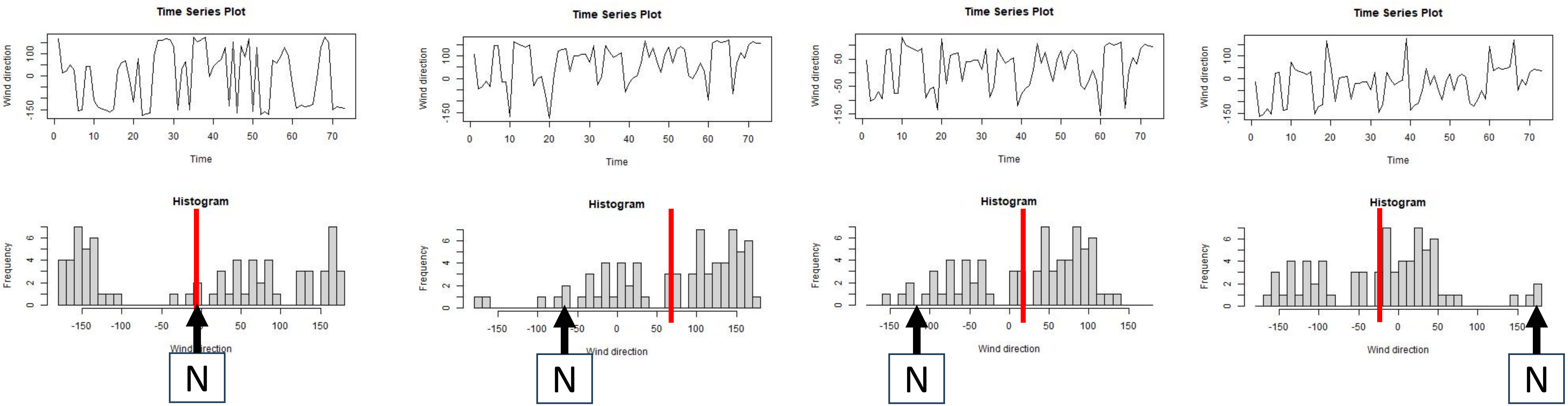
Fig. Direction of wind at pm 6:00
at Texas (2003/5/20-2003/07/31)

(a) 0 = North

(b) 0 = East-North-East

(c) 0 = South-West-South

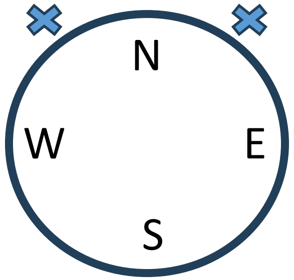
(c) 0 = South



In directional statistics, the arithmetic mean cannot become the representative value of data. For example, the observations

$$10^\circ \text{ and } 350^\circ \text{ from North} \Leftrightarrow 170^\circ \text{ and } 190^\circ \text{ from South}$$

$$\text{Mean} = 180^\circ \text{ from North} \not\Leftrightarrow \text{Mean} = 180^\circ \text{ from South}$$



Graphical representations for circular data

Fig. Direction of wind at pm 6:00
at Texas (2003/5/20-2003/07/31)

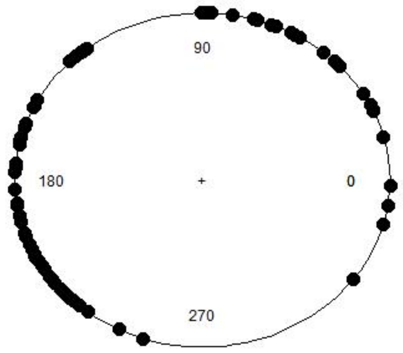
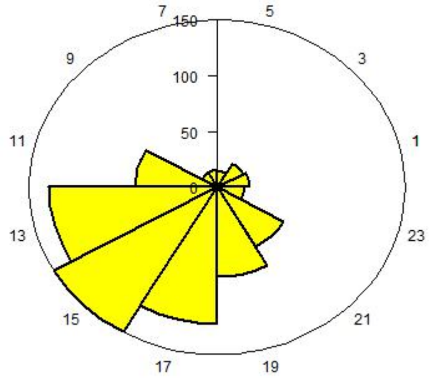


Fig. Time (by o'clock) of thunder heard
in a day at Kew Gardens (1910-1935)



Graphical representations for circular data

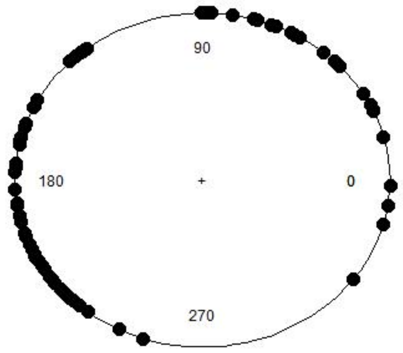


Fig. Direction of wind at pm 6:00
at Texas (2003/5/20-2003/07/31)

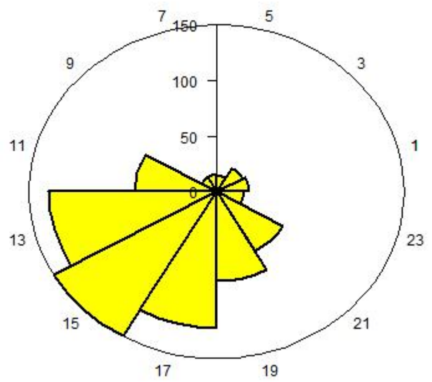
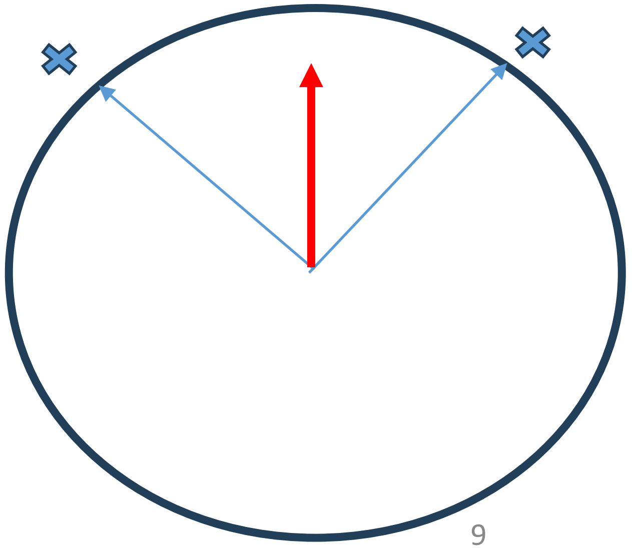


Fig. Time (by o'clock) of thunder heard
in a day at Kew Gardens (1910-1935)

Representative values for circular data

For data $\theta_1, \theta_2, \ldots, \theta_n$, put $\bar{C} = \frac{1}{n}\sum_j \cos\theta_j$, $\bar{S} = \frac{1}{n}\sum_j \sin\theta_j$



Mean direction: $\mathrm{atan2}(\bar{C}, \bar{S})$

Mean resultant length: $\sqrt{\bar{C}^2 + \bar{S}^2}$

Graphical representations for circular data

Fig. Direction of wind at pm 6:00
at Texas (2003/5/20-2003/07/31)

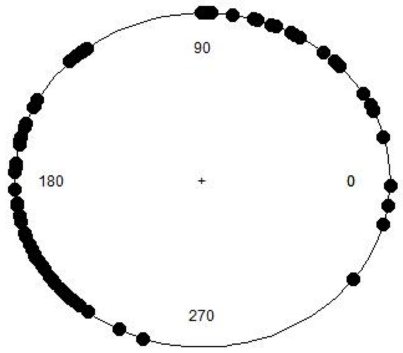
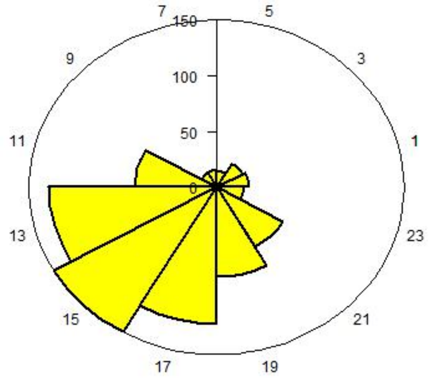


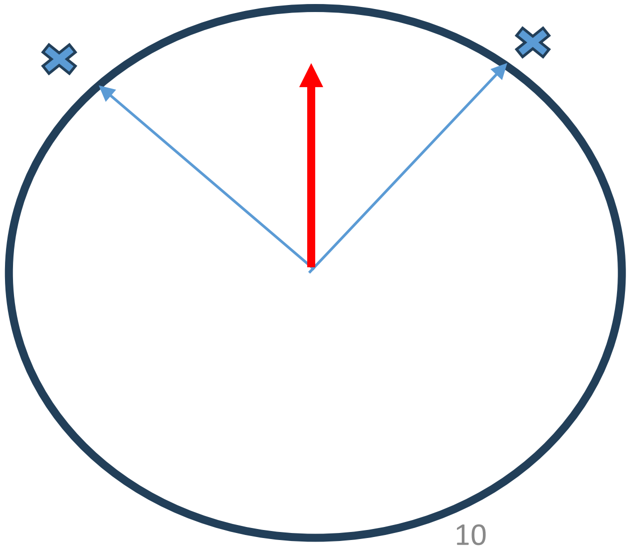
Fig. Time (by o'clock) of thunder heard
in a day at Kew Gardens (1910-1935)



Representative values for circular data

For data $\theta_1, \theta_2, \ldots, \theta_n$, put $\bar{C} = \frac{1}{n}\sum_j \cos\theta_j, \bar{S} = \frac{1}{n}\sum_j \sin\theta_j$

or $\bar{E} = \frac{1}{n}\sum_j e^{i\theta_j}$



Mean direction: $\mathrm{atan2}(\bar{C}, \bar{S})$ or $\arg(\bar{E})$

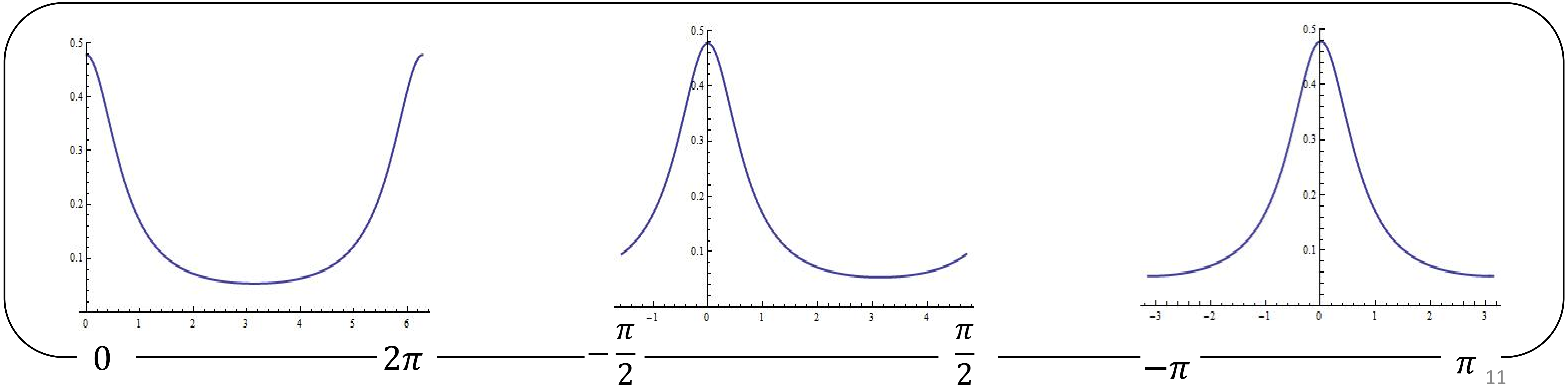
Mean resultant length: $\sqrt{\bar{C}^2 + \bar{S}^2}$ or $|\bar{E}|$

Circular distribution

The probabilistic model for circular data, which satisfies:

circular pdf $f(\theta)$

- $f(\theta) \geq 0$
- $\int_{-\pi}^{\pi} f(\theta) d\theta = 1$
- $f(\theta) = f(\theta + 2k\pi), \qquad k \in \mathbb{Z}$



Circular distribution

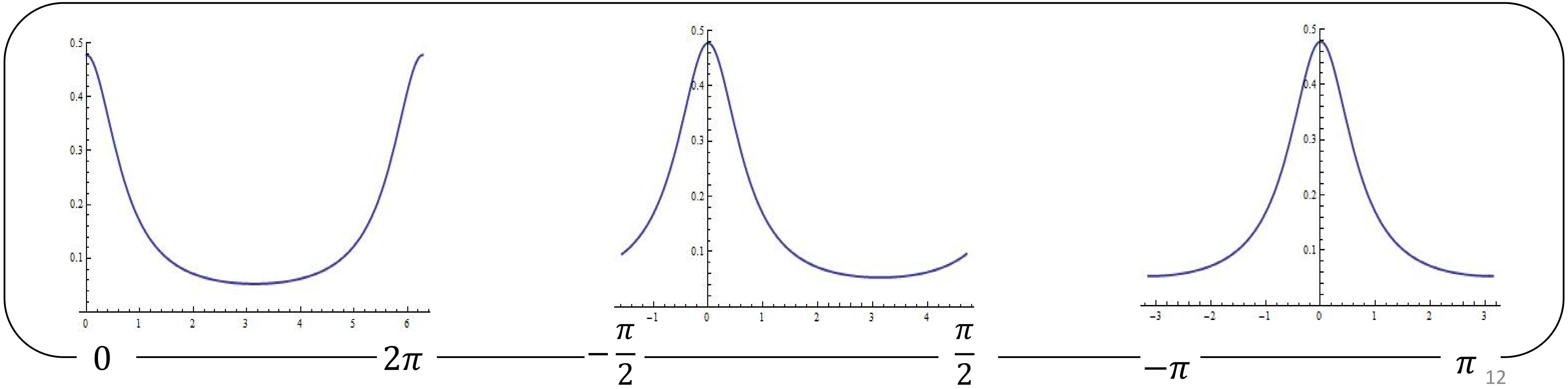
The probabilistic model for circular data, which satisfies:

circular pdf $f(\theta)$

- $f(\theta) \geq 0$
- $\int_{-\pi}^{\pi} f(\theta) d\theta = 1$
- $f(\theta) = f(\theta + 2k\pi), \qquad k \in \mathbb{Z}$

cf $\Phi_q := \int_{-\pi}^{\pi} e^{iq\Theta} f(\theta) d\theta$

- Mean direction: $\arg\{\Phi_1\}$
- Mean resultant length: $|\Phi_1|$

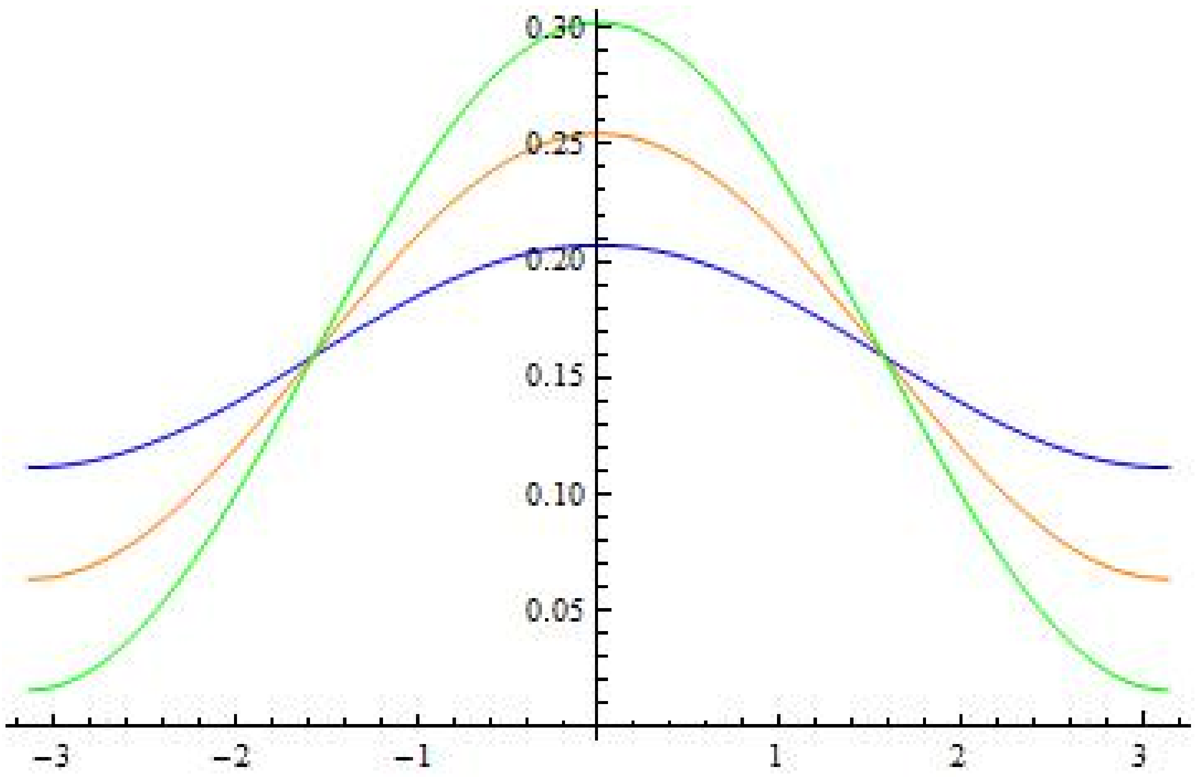


Example

Cardioid distribution

$$f(\theta) = \frac{1 - 2\rho\cos(\theta - \mu)}{2\pi}$$

- Mean direction: μ
- Mean resultant length: ρ



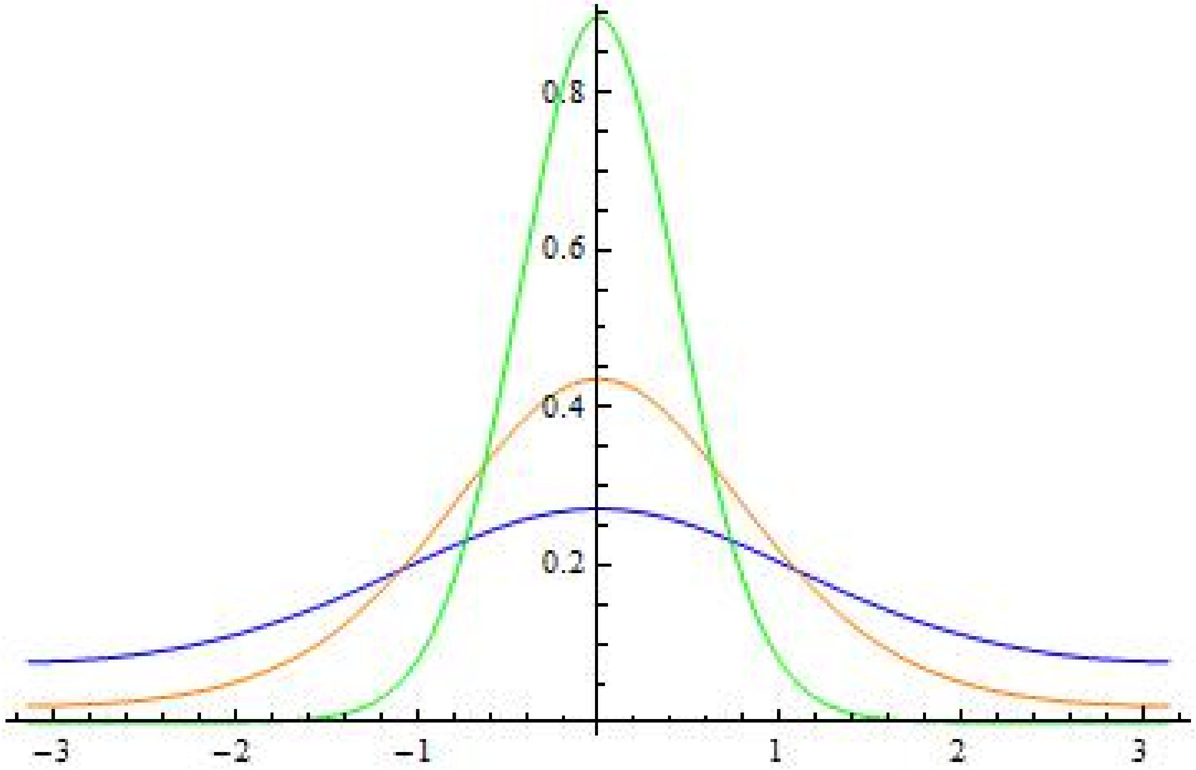
$\mu = 0$,
$\rho = 0.15$ (blue);
$\rho = 0.30$ (orange);
$\rho = 0.45$ (green);

Example

Von Mises distribution

$$f(\theta) = \frac{1}{2\pi\, I_0(\kappa)} e^{\kappa \cos(\theta-\mu)}$$

- Mean direction: μ
- Mean resultant length: $\frac{I_1(\kappa)}{I_0(\kappa)}$



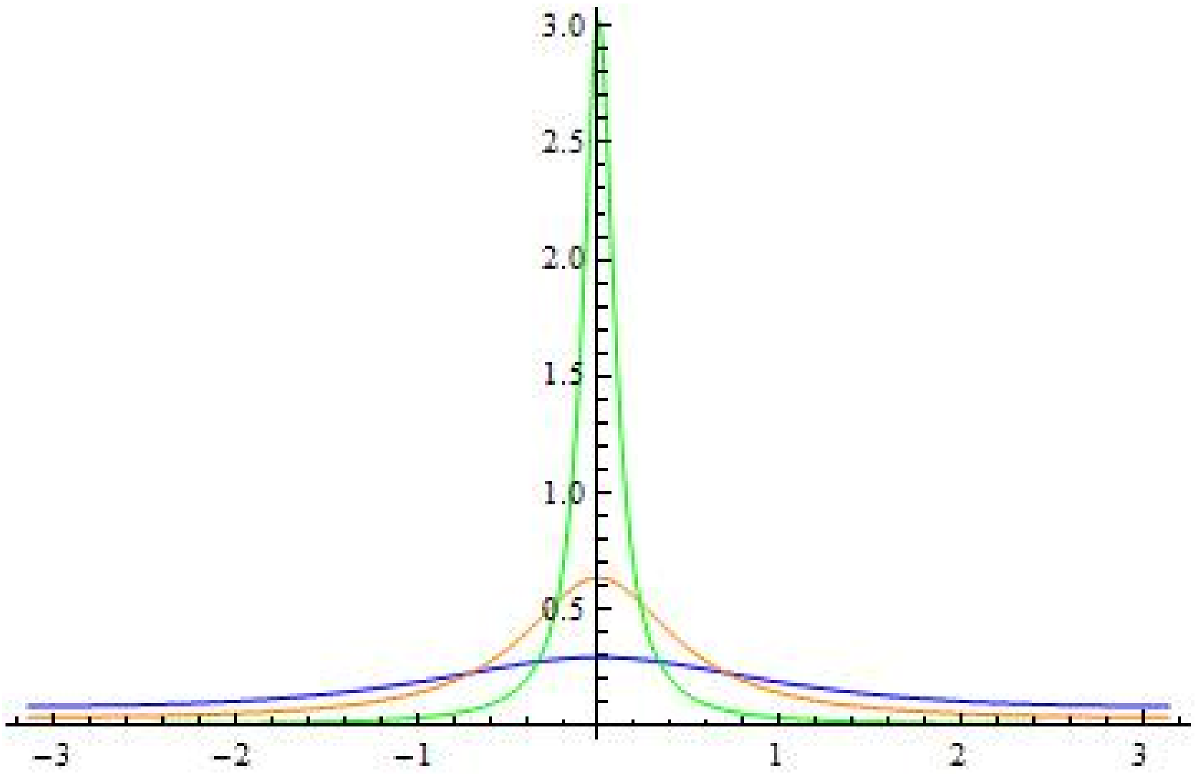
$\mu = 0,$
$\kappa = 0.6$ (blue);
$\kappa = 1.5$ (orange);
$\kappa = 5.3$ (green);

Example

Wrapped Cauchy distribution

$$f(\theta) = \frac{1-\rho^2}{2\pi\{1+\rho^2-2\rho\cos(\theta-\mu)\}}$$

- Mean direction: μ
- Mean resultant length: ρ



$\mu = 0$,
$\rho = 0.3$ (blue);
$\rho = 0.6$ (orange);
$\rho = 0.9$ (green);

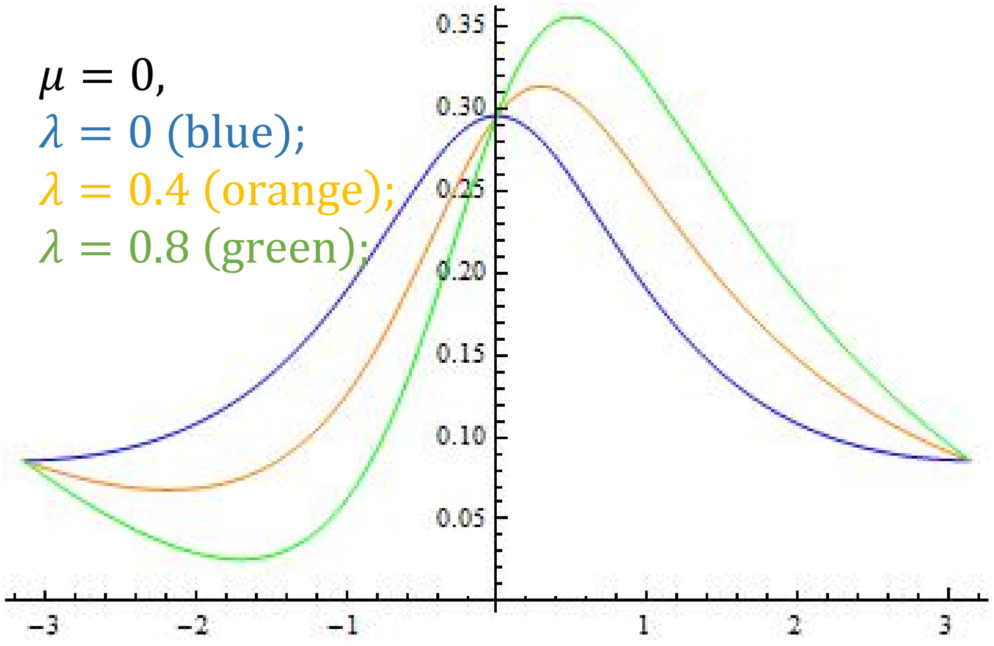
Examples of skewed distributions

$f_0(\cdot)$: Symmetric circular density about 0

Sine-skewed circular distribution

$$f(\theta) = \{1 + \lambda \sin(\theta - \mu)\} f_0(\theta - \mu)$$

by Abe and Pewsey (2011)



Examples of skewed distributions

$f_0(\cdot)$: Symmetric circular density about 0

$G(\cdot)$: Distribution function of the symmetric distribution about 0

Sine-skewed circular distribution

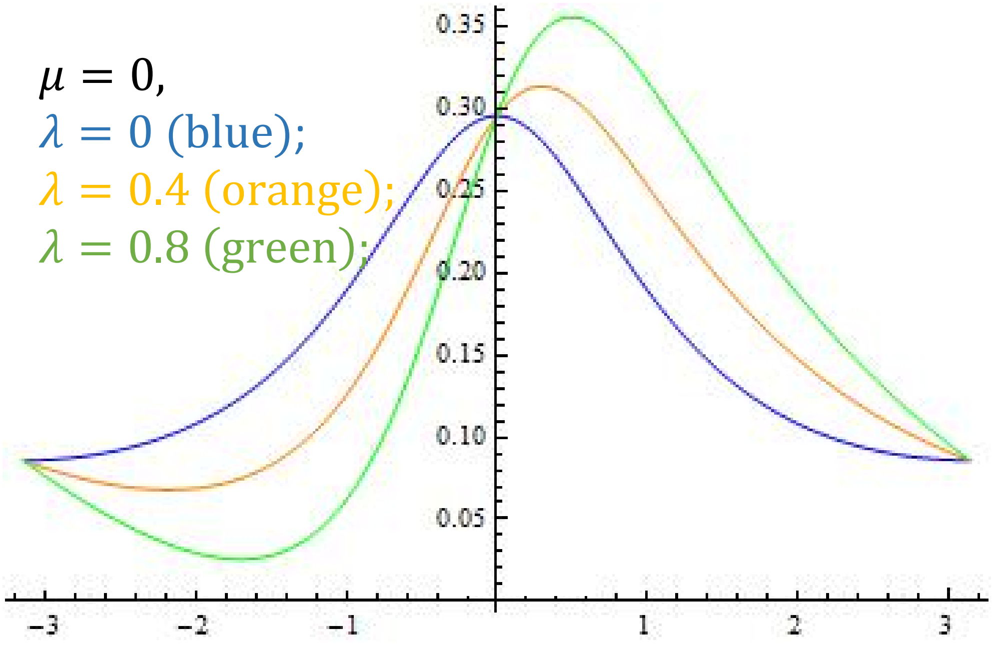
$$f(\theta) = \{1 + \lambda \sin(\theta - \mu)\} f_0(\theta - \mu)$$

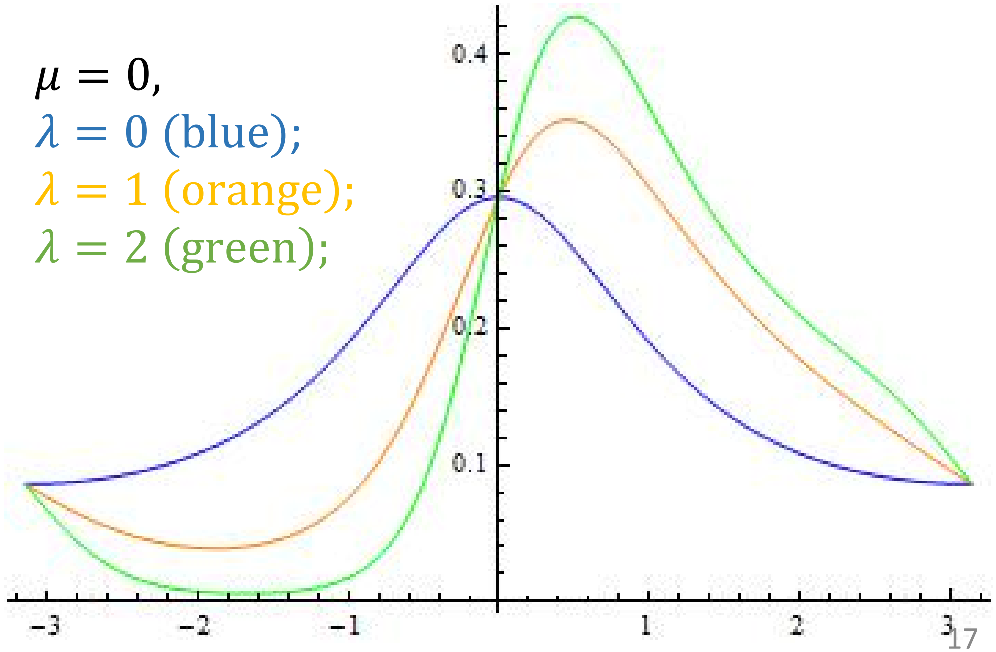
by Abe and Pewsey (2011)

More skewed circular distribution

$$f(\theta) = G(\lambda \sin(\theta - \mu)) f_0(\theta - \mu)$$

by Abe, Imoto Miyata and Shiohama (2022)

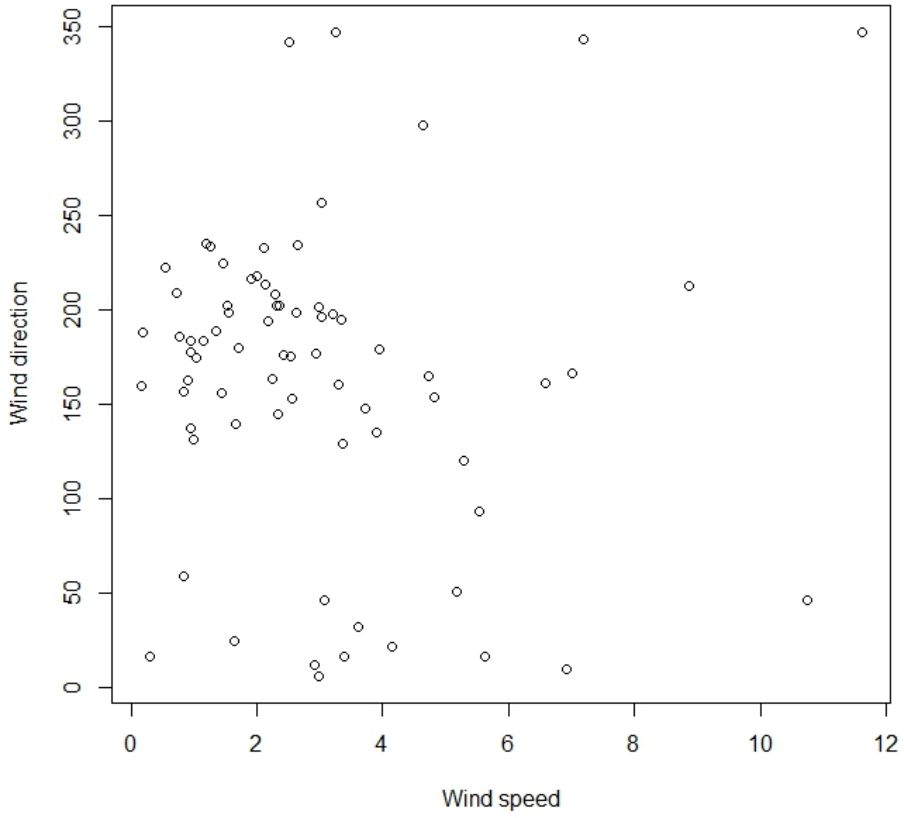


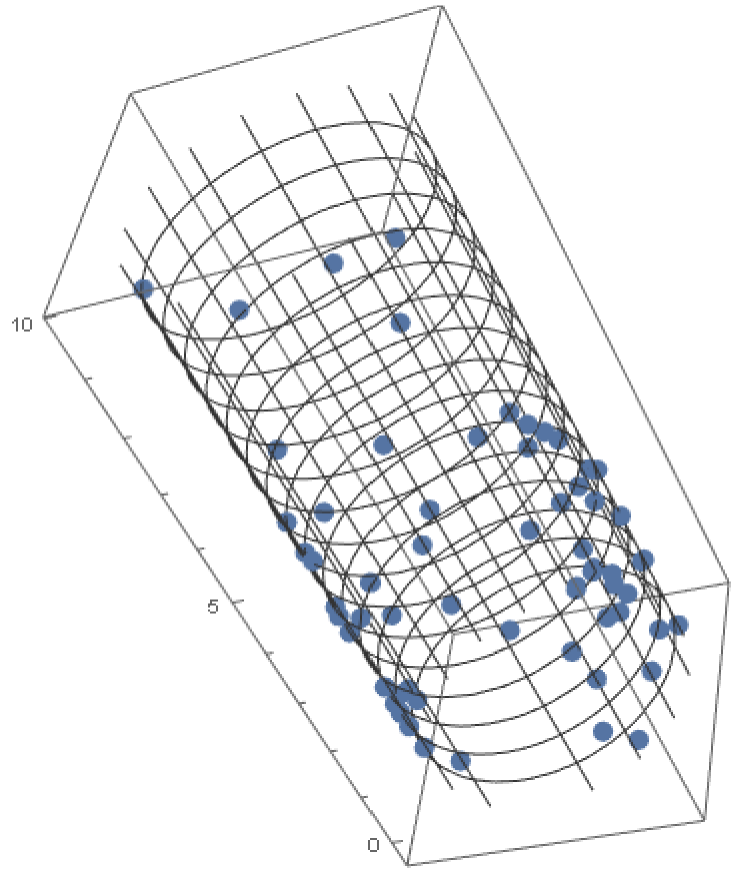


Cylindrical distribution

Cylindrical data consists of a combination of linear and circular observations,
such as the directions of wind at two points.

Fig. The speed and direction of wind at pm 6:00 at Texas (2003/5/20-2003/07/31)





Cylindrical distribution

The method for constructing cylindrical distribution by Johnson and Wehrly (1977);

$f_x(\cdot)$: linear density, $F_X(\cdot)$: corresponding distributon function

$f_\Theta(\cdot)$: circular density, $F_\Theta(\cdot)$: corresponding distributon function

$$f(x,\theta) = 2\pi\ g(2\pi\{F_X(x) + pF_\Theta(\theta)\})\ f_X(x)\ f_\Theta(\theta),$$

where $p \in \{-1,1\}$, and $g(\cdot)$ is circular density.

$\Longrightarrow$ The marginal densities are $f_X(x)$ and $f_\Theta(\theta)$.

Cylindrical distribution

The method for constructing cylindrical distribution by Johnson and Wehrly (1977);

$f_x(\cdot)$: linear density, $F_X(\cdot)$: corresponding distributon function

$f_\Theta(\cdot)$: circular density, $F_\Theta(\cdot)$: corresponding distributon function

$$f(x,\theta) = 2\pi\ g(2\pi\{F_X(x) + pF_\Theta(\theta)\})\ f_X(x)\ f_\Theta(\theta),$$

where $p \in \{-1,1\}$, and $g(\cdot)$ is circular density.

$\Longrightarrow$ The marginal densities are $f_X(x)$ and $f_\Theta(\theta)$.

This construction needs the calculation of the distribution functions.

$\Longrightarrow$ Propose the simple construction that need not complicated calculations.

Proposition

New distribution family

$G(\cdot)$: distribution function of the symmetric distribution, i.e., $G(x) = 1 - G(-x)$

Examples

- $G(x) = \frac{1+x}{2},\ x \in [-1,1]$: Uniform df on $[-1,1]$
- $G(x) = \frac{1}{1+e^{-x}},\ x \in \mathbb{R}$: Logistic df
- $G(x) = \frac{1}{2}\left\{1 + \frac{2}{\pi}\mathrm{Arctan}(x)\right\},\ x \in \mathbb{R}$: Cauchy df
- $G(x) = \frac{1}{2}\left\{1 + \frac{2}{\pi}\mathrm{Arctan}(e^{\kappa}\tan x)\right\},\ x \in [-\pi,\pi)$:Wrapped Cauchy df

Proposition

New distribution family

$G(\cdot)$: distribution function of the symmetric distribution, i.e., $G(x) = 1 - G(-x)$

Examples

- $G(x) = \frac{1+x}{2},\ x \in [-1,1]$: Uniform df on $[-1,1]$
- $G(x) = \frac{1}{1+e^{-x}},\ x \in \mathbb{R}$: Logistic df
- $G(x) = \frac{1}{2}\left\{1 + \frac{2}{\pi}\mathrm{Arctan}(x)\right\},\ x \in \mathbb{R}$: Cauchy df
- $G(x) = \frac{1}{2}\left\{1 + \frac{2}{\pi}\mathrm{Arctan}(e^{\kappa}\tan x)\right\},\ x \in [-\pi, \pi)$:Wrapped Cauchy df

$w(\cdot,\cdot)$: $w(x,\theta) = -w(-x,\theta) = -w(x,\theta) = w(-x,-\theta)$,
$w(x,\theta) = w(x,\theta + 2\pi)$ for $k \in \mathbb{Z}$, and $w(x,\theta) \in \mathrm{supp}\{G\}$

Examples

- $w(x,\theta) = x \sin\theta$
- $w(x,\theta) = \sin(2\mathrm{Arctan}\ x)\sin\theta = \frac{2x}{1+x^2}\sin\theta$

$f_X(\cdot)$: symmetric linear density whose mean is 0 and variance is 1

$f_\Theta(\cdot)$: symmetric circular density whose mean direction is 0

Proposed distribution family

$$f(x,\theta) = 2G\left(\lambda\, w\left(\frac{x-\mu_X}{\tau}, \theta-\mu_\Theta\right)\right)\frac{1}{\sigma} f_X\left(\frac{x-\mu_X}{\sigma}\right) f_\Theta(\theta-\mu_\Theta)$$

Hereafter, (X, Θ) is assumed to be the rv of the above distribution.

λ and τ: correlation parameters between X and Θ

μ_X : location parameter of X

σ : scale parameter of X

μ_Θ : location parameter of Θ

Property

$$f(x,\theta) = 2G\left(\lambda\, w\left(\frac{x-\mu_X}{\tau}, \theta-\mu_\Theta\right)\right)\frac{1}{\sigma} f_X\left(\frac{x-\mu_X}{\sigma}\right) f_\Theta(\theta-\mu_\Theta)$$

Marginal distribution

The marginal densities of X and Θ are

$$\frac{1}{\sigma} f_X\left(\frac{x-\mu_X}{\sigma}\right) \quad \text{and} \quad f_\Theta(\theta-\mu_\Theta),$$

respectively.

The copula for the proposed distribution (when $\mu_X = \mu_\Theta = 0, \sigma = 1$) is

$$c(x,\theta) = G\left(\lambda\, w\left(\frac{F_X^{-1}(x)}{\tau}, F_\Theta^{-1}(\theta)\right)\right).$$

Property

$$f(x,\theta) = 2G\left(\lambda\, w\left(\frac{x-\mu_X}{\tau}, \theta-\mu_\Theta\right)\right)\frac{1}{\sigma} f_X\left(\frac{x-\mu_X}{\sigma}\right) f_\Theta(\theta-\mu_\Theta)$$

Conditional distribution

The conditional density of X given $\Theta = \theta$ is

$$2G\left(\lambda\, w\left(\frac{x-\mu_X}{\tau}, \theta-\mu_\Theta\right)\right)\frac{1}{\sigma} f_X\left(\frac{x-\mu_X}{\sigma}\right), \qquad \text{and}$$

The conditional density of Θ given $\mathrm{X} = x$ is

$$2G\left(\lambda\, w\left(\frac{x-\mu_X}{\tau}, \theta-\mu_\Theta\right)\right) f_\Theta(\theta-\mu_\Theta)$$

corresponds to the distribution family
by Abe, Imoto, Miyata and Shiohama (2022)

Property

$$f(x,\theta) = 2G\left(\lambda\, w\left(\frac{x-\mu_X}{\tau}, \theta - \mu_\Theta\right)\right)\frac{1}{\sigma} f_X\left(\frac{x-\mu_X}{\sigma}\right) f_\Theta(\theta - \mu_\Theta)$$

Random number generation

- Generate three random numbers by $z \sim G$, y $\sim f_X, \psi \sim f_\Theta$.

- Put

$$(x,\theta) = \begin{cases} (y,\psi) & if\ z < \lambda\, w(y,\psi) \\ (-y,\psi) & if\ z \geq \lambda w(y,\psi) \end{cases}$$

Then, (x,θ) is the random number of the proposed distribution with $\mu_X = \mu_\Theta = 0, \sigma = 1$.

Property

$$f(x,\theta) = 2G\left(\lambda\, w\left(\frac{x-\mu_X}{\tau}, \theta-\mu_\Theta\right)\right)\frac{1}{\sigma}f_X\left(\frac{x-\mu_X}{\sigma}\right)f_\Theta(\theta-\mu_\Theta)$$

ML estimation

Since the log-likelihood function is

$$\log G\left(\lambda\, w\left(\frac{x-\mu_X}{\tau}, \theta-\mu_\Theta\right)\right) + \log f_X\left(\frac{x-\mu_X}{\sigma}\right) - \log\sigma + \log f_\Theta(\theta-\mu_\Theta) - \log 2,$$

the parameters in $f_X(\cdot)$ and $f_\Theta(\cdot)$, except μ_X and μ_Θ, can be estimated independently.

Property

$$f(x,\theta) = 2G\left(\lambda\, w\left(\frac{x-\mu_X}{\tau}, \theta-\mu_\Theta\right)\right)\frac{1}{\sigma} f_X\left(\frac{x-\mu_X}{\sigma}\right) f_\Theta(\theta-\mu_\Theta)$$

Fisher Information

Put $\iota_{\alpha\beta} = -\mathrm{E}\left[\frac{\partial \log f}{\partial\alpha\partial\beta}\right]$. Then Fisher information matrix contains many 0, or

$$\begin{matrix} \text{About } \lambda \longrightarrow \\ \text{About } \tau \longrightarrow \\ \text{About } \mu_X \longrightarrow \\ \text{About } \sigma \longrightarrow \\ \text{About } \mu_\Theta \longrightarrow \end{matrix} \begin{pmatrix} \iota_{\lambda\lambda} & \iota_{\lambda\tau} & 0 & 0 & 0 \\ \iota_{\lambda\tau} & \iota_{\tau\tau} & \iota_{\tau\mu_X} & 0 & 0 \\ 0 & \iota_{\tau\mu_X} & \iota_{\mu_X\mu_X} & \iota_{\mu_X\sigma} & 0 \\ 0 & 0 & \iota_{\mu_X\sigma} & \iota_{\sigma\sigma} & 0 \\ 0 & 0 & 0 & 0 & \iota_{\mu_\Theta\mu_\Theta} \end{pmatrix}$$

Special case

$$f(x,\theta) = \left[1 + 2\lambda \frac{\left(\frac{x-\mu_X}{\tau}\right)\sin(\theta-\mu_\Theta)}{1+\left(\frac{x-\mu_X}{\tau}\right)^2}\right] \frac{1}{\sigma} f_X\left(\frac{x-\mu_X}{\sigma}\right) f_\Theta(\theta-\mu_\Theta)$$

Special case

$$f(x,\theta) = \left[1 + 2\lambda \frac{\left(\frac{x-\mu_X}{\tau}\right)\sin(\theta-\mu_\Theta)}{1+\left(\frac{x-\mu_X}{\tau}\right)^2}\right] \frac{1}{\sigma} f_X\left(\frac{x-\mu_X}{\sigma}\right) f_\Theta(\theta-\mu_\Theta)$$

Put $\gamma_p := \mathrm{E}\left[\left(\frac{X-\mu_X}{\sigma}\right)^p\right]$ and $\alpha_q + i\beta_q := \mathrm{E}\left[e^{iq(\Theta-\mu_\Theta)}\right]$. (When Θ is symmetric, $\beta_q = 0$)

Then the joint moment is

$$\mathrm{E}\left[\left(\frac{X-\mu_X}{\sigma}\right)^p e^{iq(\Theta-\mu_\Theta)}\right] = \begin{cases} \gamma_p(\alpha_q + i\beta_q) & p: \text{even} \\ \lambda A_p(\tau,\sigma)\left\{\left(\beta_{q+1} - \beta_{q-1}\right) - i\left(\alpha_{q+1} - i\alpha_{q-1}\right)\right\} & p: \text{odd} \end{cases},$$

where $A_p(\tau,\sigma) = \int \frac{\sigma\, x^{p+1}/\tau}{1+(\sigma x/\tau)^2} f_X(x)dx$.

Special case

$$f(x,\theta) = \left[1 + 2\lambda \frac{\left(\frac{x-\mu_X}{\tau}\right)\sin(\theta-\mu_\Theta)}{1+\left(\frac{x-\mu_X}{\tau}\right)^2}\right]\frac{1}{\sigma} f_X\left(\frac{x-\mu_X}{\sigma}\right) f_\Theta(\theta-\mu_\Theta)$$

Put $\gamma_p := \mathrm{E}\left[\left(\frac{X-\mu_X}{\sigma}\right)^p\right]$ and $\alpha_q + i\beta_q := \mathrm{E}\left[e^{iq(\Theta-\mu_\Theta)}\right]$. (When Θ is symmetric, $\beta_q = 0$)

Then the joint moment is

$$\mathrm{E}\left[\left(\frac{X-\mu_X}{\sigma}\right)^p e^{iq(\Theta-\mu_\Theta)}\right] = \begin{cases} \gamma_p(\alpha_q + i\beta_q) & p\text{: even} \\ \lambda A_p(\tau,\sigma)\{(\beta_{q+1}-\beta_{q-1}) - i(\alpha_{q+1} - i\alpha_{q-1})\} & p\text{: odd} \end{cases},$$

where $A_p(\tau,\sigma) = \int \frac{\sigma\, x^{p+1}/\tau}{1+(\sigma x/\tau)^2} f_X(x)dx$.

The cylindrical correlation by Mardia (1976) and Johnson and Wehrly (1977) is

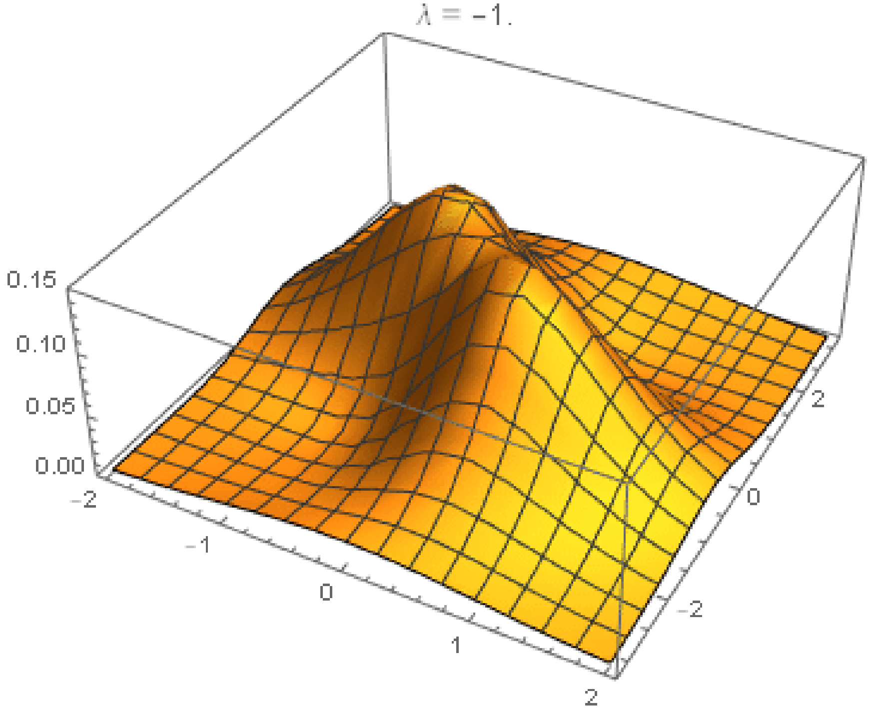
$$R_{X\Theta}^2 = \frac{\mathrm{Corr}[X,\cos\Theta]^2 + \mathrm{Corr}[X,\sin\Theta]^2 - 2\mathrm{Corr}[\cos\Theta,\sin\Theta]\mathrm{Corr}[X,\cos\Theta]\mathrm{Corr}[X,\sin\Theta]}{1-\mathrm{Corr}[\cos\Theta,\sin\Theta]^2}$$

$$= 2\lambda^2 A_1^2(\tau,\sigma)(1-\alpha_2)$$

Special case

$$f(x,\theta) = \left[1 + 2\lambda \frac{\left(\frac{x-\mu_X}{\tau}\right)\sin(\theta-\mu_\Theta)}{1+\left(\frac{x-\mu_X}{\tau}\right)^2}\right] \frac{1}{\sigma} f_X\left(\frac{x-\mu_X}{\sigma}\right) f_\Theta(\theta-\mu_\Theta)$$

When $f_X(x) = \frac{1}{2\pi} e^{-\frac{x^2}{2}}$, $f_\Theta(\theta) = \frac{e^{\cos\theta}}{2\pi\, I_0(1)}$, and $\mu_X = \mu_\Theta = 0, \sigma = 1$,

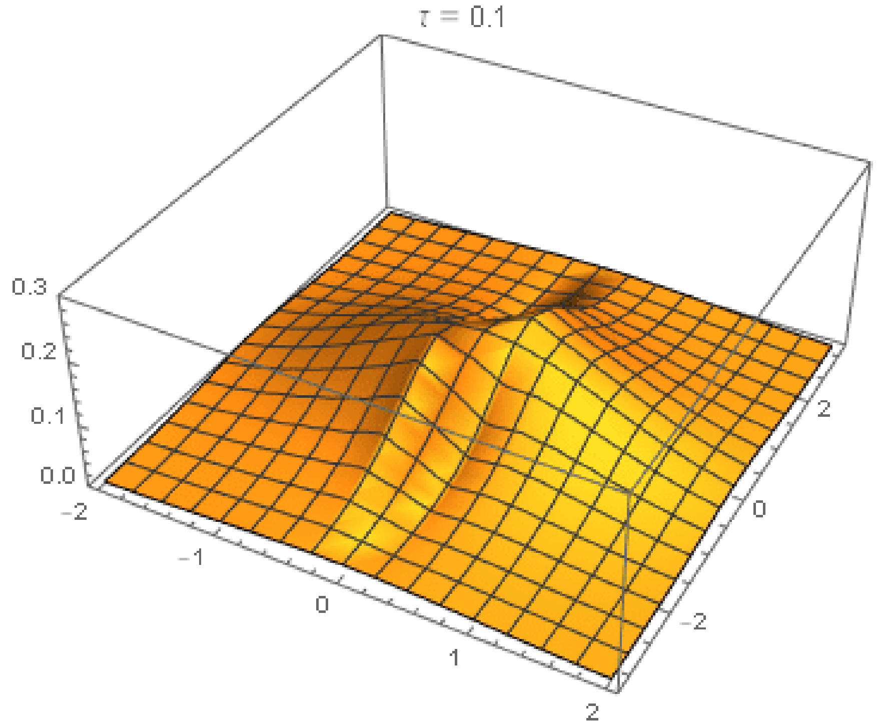


← Plot with different λ

Special case

$$f(x,\theta) = \left[1 + 2\lambda \frac{\left(\frac{x-\mu_X}{\tau}\right)\sin(\theta-\mu_\Theta)}{1+\left(\frac{x-\mu_X}{\tau}\right)^2}\right]\frac{1}{\sigma} f_X\left(\frac{x-\mu_X}{\sigma}\right) f_\Theta(\theta-\mu_\Theta)$$

When $f_X(x) = \frac{1}{2\pi} e^{-\frac{x^2}{2}}$, $f_\Theta(\theta) = \frac{e^{\cos\theta}}{2\pi\, I_0(1)}$, and $\mu_X = \mu_\Theta = 0, \sigma = 1$,



← Plot with different τ

Special case

$$f(x,\theta) = \left[1 + 2\lambda \frac{\left(\frac{x-\mu_X}{\tau}\right)\sin(\theta-\mu_\Theta)}{1+\left(\frac{x-\mu_X}{\tau}\right)^2}\right]\frac{1}{\sigma} f_X\left(\frac{x-\mu_X}{\sigma}\right) f_\Theta(\theta-\mu_\Theta)$$

Put $\gamma_p := \mathrm{E}\left[\left(\frac{X-\mu_X}{\sigma}\right)^p\right]$ and $\alpha_q + i\beta_q := \mathrm{E}\left[e^{iq(\Theta-\mu_\Theta)}\right]$. (When Θ is symmetric, $\beta_q = 0$)

The conditional distribution of Θ belongs to the sine-skewed circular distribution by Abe and Pewsey (2011)

$$f(\theta \mid x) = \left[1 + 2\lambda \frac{\left(\frac{x-\mu_X}{\tau}\right)\sin(\theta-\mu_\Theta)}{1+\left(\frac{x-\mu_X}{\tau}\right)^2}\right] f_\Theta(\theta-\mu_\Theta),$$

and the conditional moment is given by

$$\mathrm{E}\left[e^{iq(\Theta-\mu_\Theta)} \mid X = x\right]$$

$$= \alpha_q + \lambda(\beta_{q+1} - \beta_{q-1})\sin\left(2\mathrm{Arctan}\left(\frac{x-\mu_X}{\sigma}\right)\right) + i\left\{\beta_q - \lambda(\alpha_{q+1} - \alpha_{q-1})\sin\left(2\mathrm{Arctan}\left(\frac{x-\mu_X}{\sigma}\right)\right)\right\}.$$

Special case

$$f(x,\theta) = \left[1 + 2\lambda \frac{\left(\frac{x-\mu_X}{\tau}\right)\sin(\theta-\mu_\Theta)}{1+\left(\frac{x-\mu_X}{\tau}\right)^2}\right]\frac{1}{\sigma} f_X\left(\frac{x-\mu_X}{\sigma}\right) f_\Theta(\theta-\mu_\Theta)$$

Put $\gamma_p := \mathrm{E}\left[\left(\frac{X-\mu_X}{\sigma}\right)^p\right]$ and $\alpha_q + i\beta_q := \mathrm{E}\left[e^{iq(\Theta-\mu_\Theta)}\right]$. (When Θ is symmetric, $\beta_q = 0$)

The conditional distribution of X is

$$f(\theta \mid x) = \left[1 + 2\lambda \frac{\left(\frac{x-\mu_X}{\tau}\right)\sin(\theta-\mu_\Theta)}{1+\left(\frac{x-\mu_X}{\tau}\right)^2}\right]\frac{1}{\sigma} f_X\left(\frac{x-\mu_X}{\sigma}\right),$$

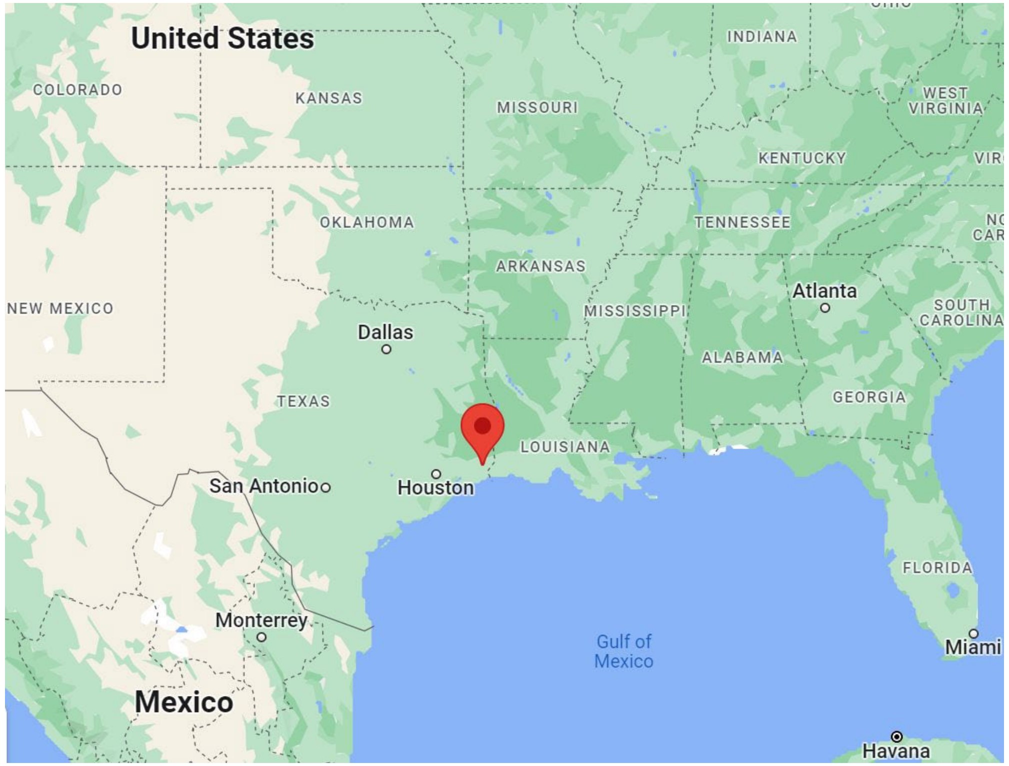
and the conditional moment is given by

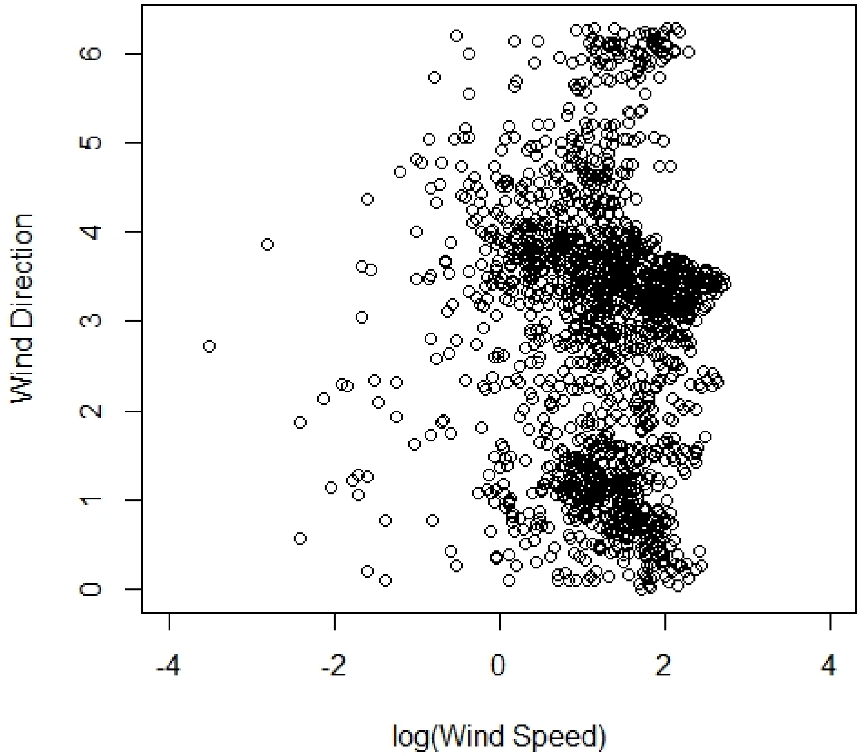
$$\mathrm{E}\left[\left(\frac{X-\mu_X}{\sigma}\right)^p \,\middle|\, \Theta = \theta\right] = \gamma_p + \lambda \sin(\theta-\mu_\Theta) A_p(\tau,\sigma)\,.$$

Illustrative application

Wind data

The speed and direction of wind per hour at Texas (2003/5/20-2003/07/31).
The sample size is 1752.





Fitting model

For the dataset (x_j, θ_j), $j = 1,2, \ldots, n$, fit the mixture model of the proposed model, i.e.,

$$h(x, \theta \mid \Xi) = \sum_{g=1}^{G} \phi_g \, f\left(x, \theta \mid \lambda_g, \tau_g, \mu_{Xg}, \sigma_g, \mu_{\Theta g}, \kappa_g\right),$$

where

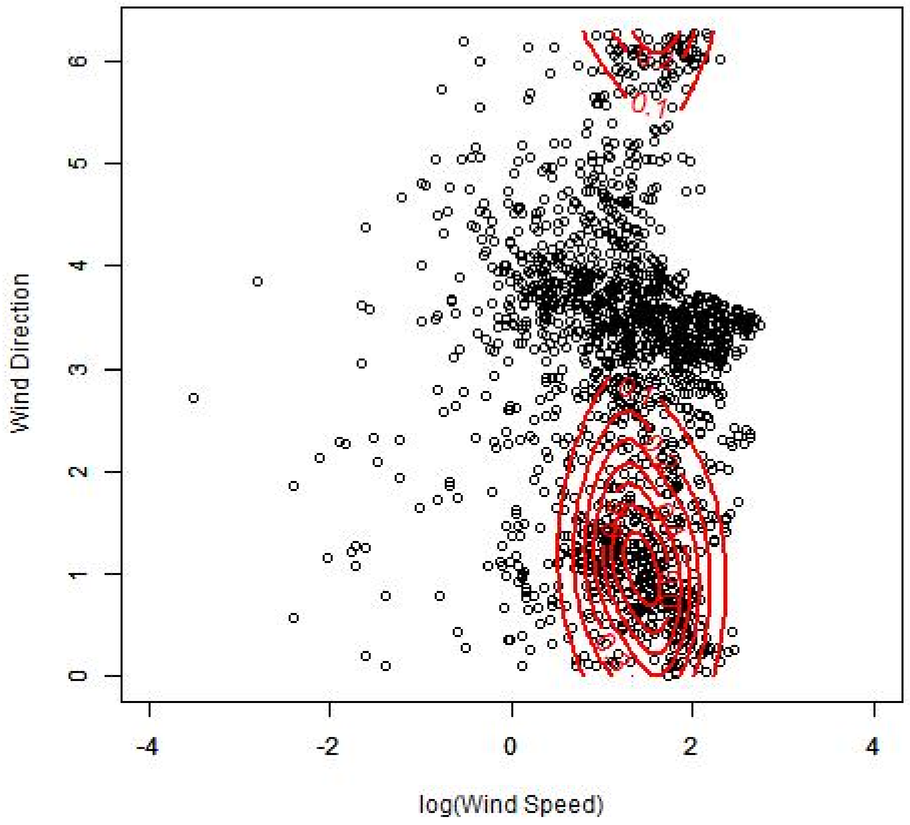
$$f(x, \theta \mid \lambda, \tau, \mu_X, \sigma, \mu_\Theta, \kappa) = \left[1 + 2\lambda \frac{\left(\frac{x - \mu_X}{\tau}\right) \sin(\theta - \mu_\Theta)}{1 + \left(\frac{x - \mu_X}{\tau}\right)^2}\right] \times N(\mu_X, \sigma^2) \times vM(\mu_\Theta, \kappa)$$

: Proposed (Normal × von Mises) model

The estimation is by the EM algorithm, and the initial values are determined by the k-means method for the transformed data $(x_j, \cos\theta_j, \sin\theta_j)$, $j = 1,2, \ldots, n$.

Result

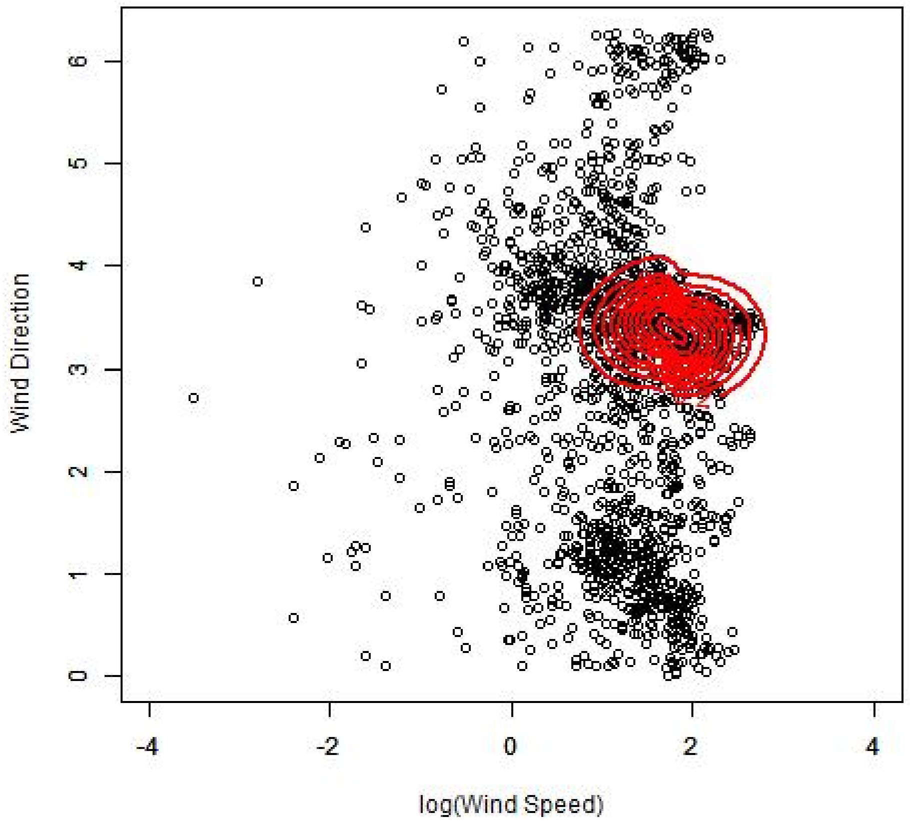
$$\phi = 0.39$$



$$\lambda = -0.31, \tau = 0.39$$
$$\mu_X = 1.44, \sigma = 0.46$$
$$\mu_\Theta = 1.05, \kappa = 1.51$$

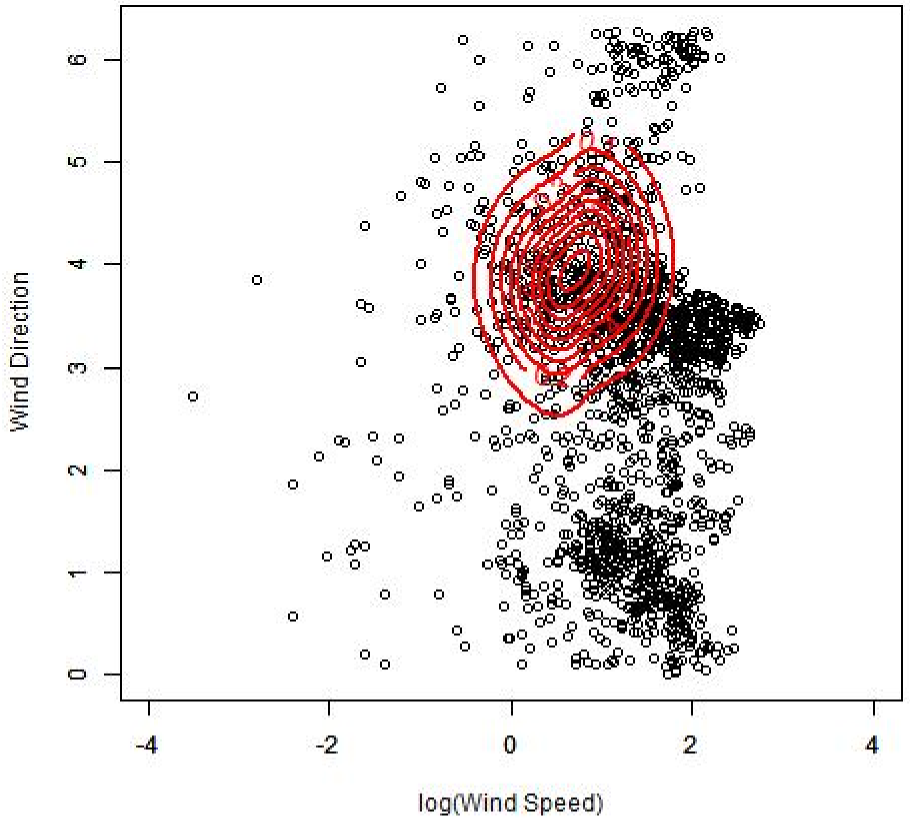
Result

$\phi = 0.29$



$\lambda = -0.97, \tau = 0.20$
$\mu_X = 1.78, \sigma = 0.47$
$\mu_\Theta = -2.93, \kappa = 11.10$

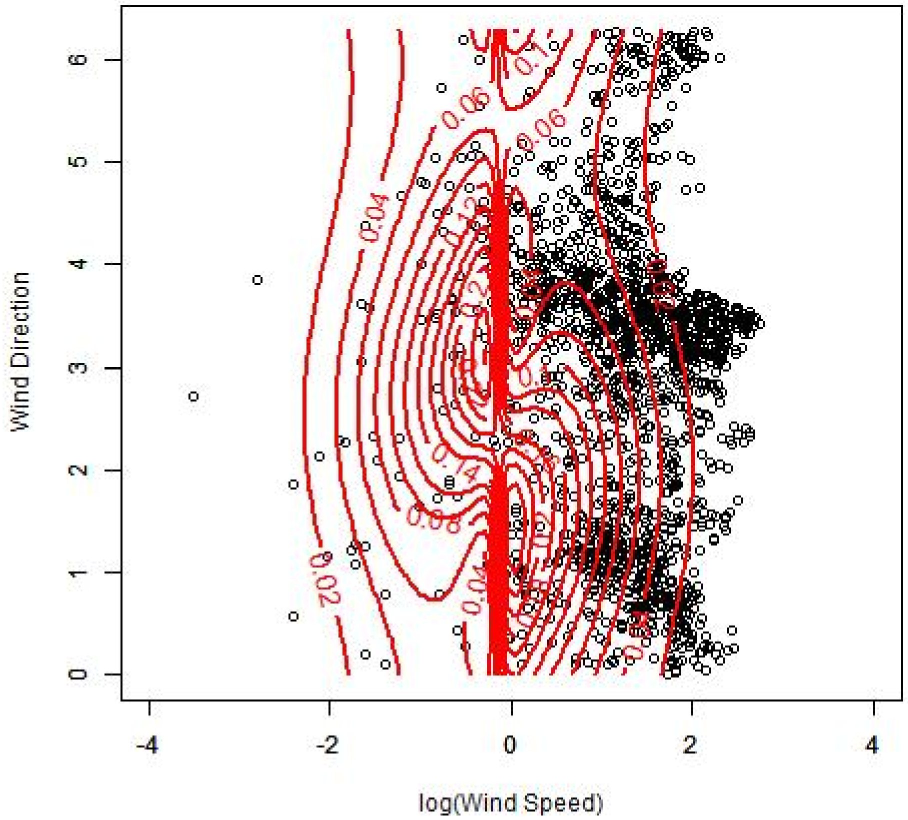
$\phi = 0.23$



$\lambda = 0.43, \tau = 0.37$
$\mu_X = 0.70, \sigma = 0.51$
$\mu_\Theta = -2.34, \kappa = 3.04$

Result

$$\phi = 0.09$$

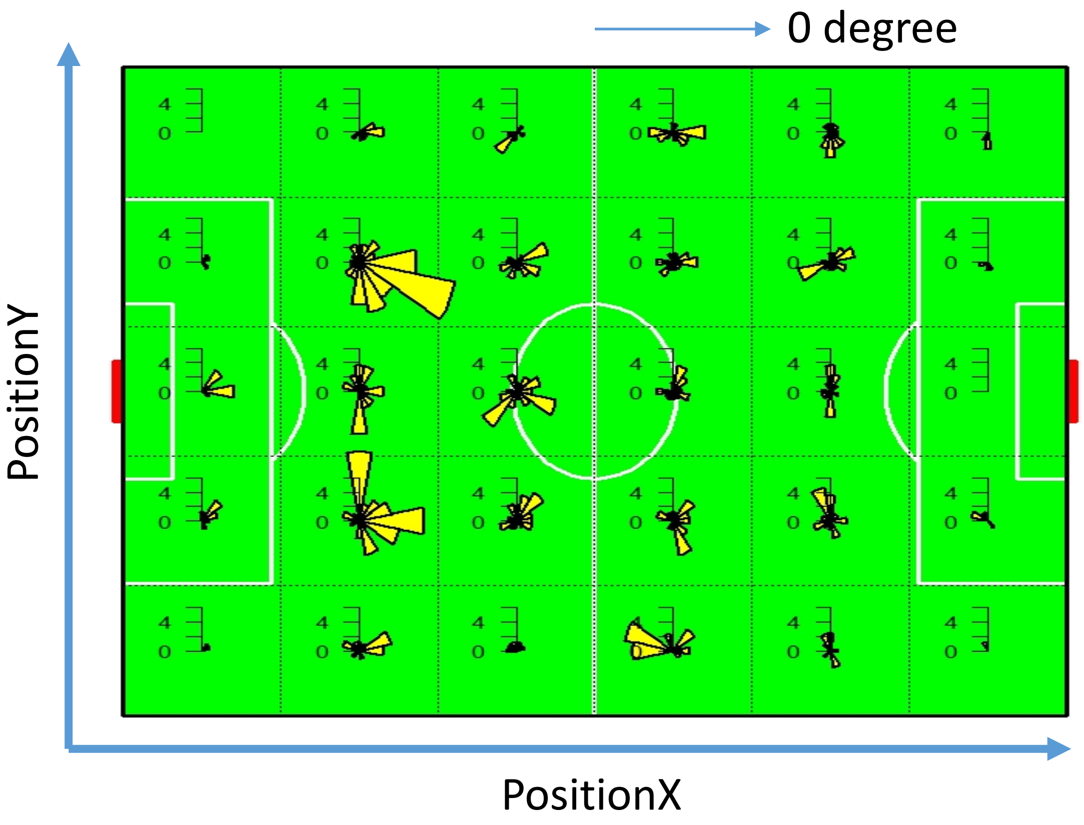


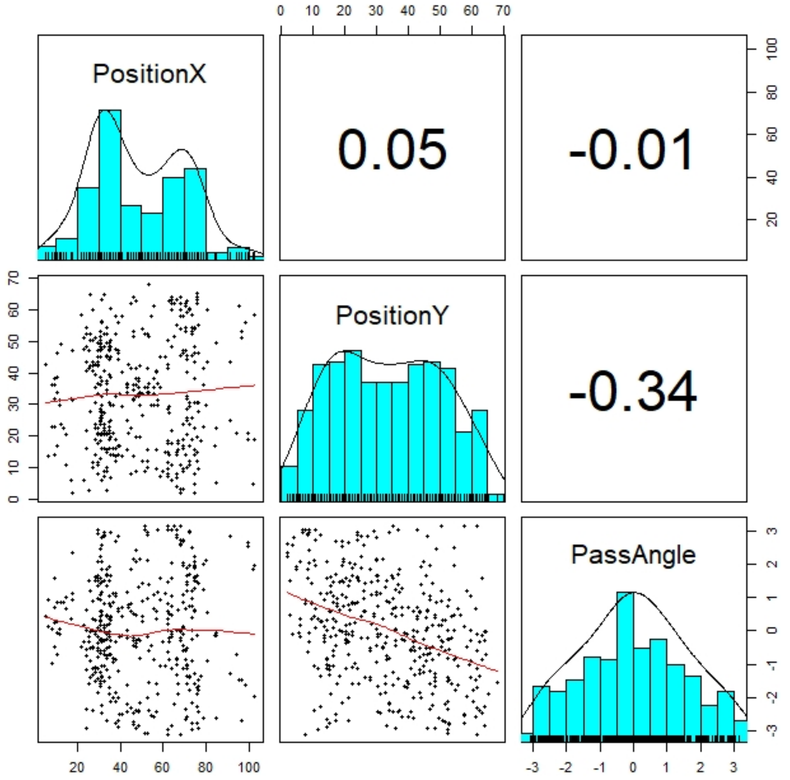
$$\lambda = -0.83, \tau = 0.16$$
$$\mu_X = -0.14, \sigma = 1.01$$
$$\mu_\Theta = 2.27, \kappa = 0.46$$

Pass sonar data

The position and direction of the ball-passing in a soccer game

Example: <u>Japan</u> - Senegal





This can be constructed from the dataset on Pappalardo, et al (2019a) and Pappalardo, et al (2019b).

Fitting model

For the dataset (x_j, y_j, θ_j), $j = 1,2,\ldots,n$, fit the mixture model of the proposed model, i.e.,

$$h(x, y, \theta \mid \Xi) = \sum_{g=1}^{G} \phi_g \, f\left(x, y, \theta \mid \lambda_g, \tau_g, \rho_g, \mu_{Xg}, \sigma_{Xg}, \mu_{Yg}, \sigma_{Yg}, \mu_{\Theta g}, \kappa_g\right),$$

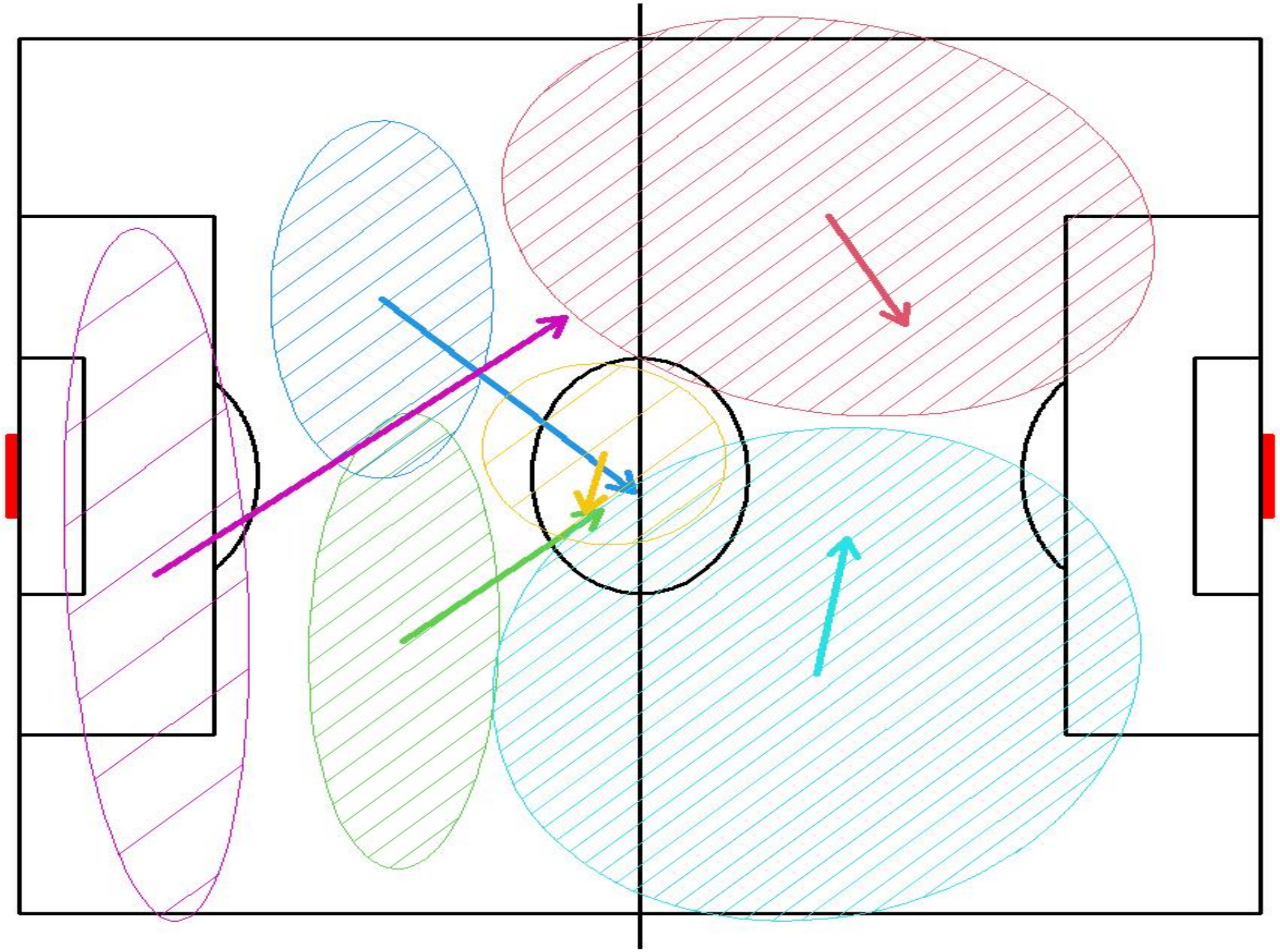
where

$$f(x, y, \theta \mid \lambda, \tau, \rho, \mu_X, \sigma_X, \mu_Y, \sigma_Y, \mu_\Theta, \kappa) = \left[1 + 2\lambda \frac{\left(\frac{y-\mu_Y}{\tau}\right)\sin(\theta - \mu_\Theta)}{1 + \left(\frac{y-\mu_Y}{\tau}\right)^2}\right] \times N_2\left(\begin{pmatrix}\mu_X \\ \mu_Y\end{pmatrix}, \begin{pmatrix}\sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2\end{pmatrix}\right) \times vM(\mu_\Theta, \kappa)$$

: Proposed (Bivariate normal × von Mises) model

Result

Japan – Senegal
in FIFA World cup 2018
(The sample size is 425)



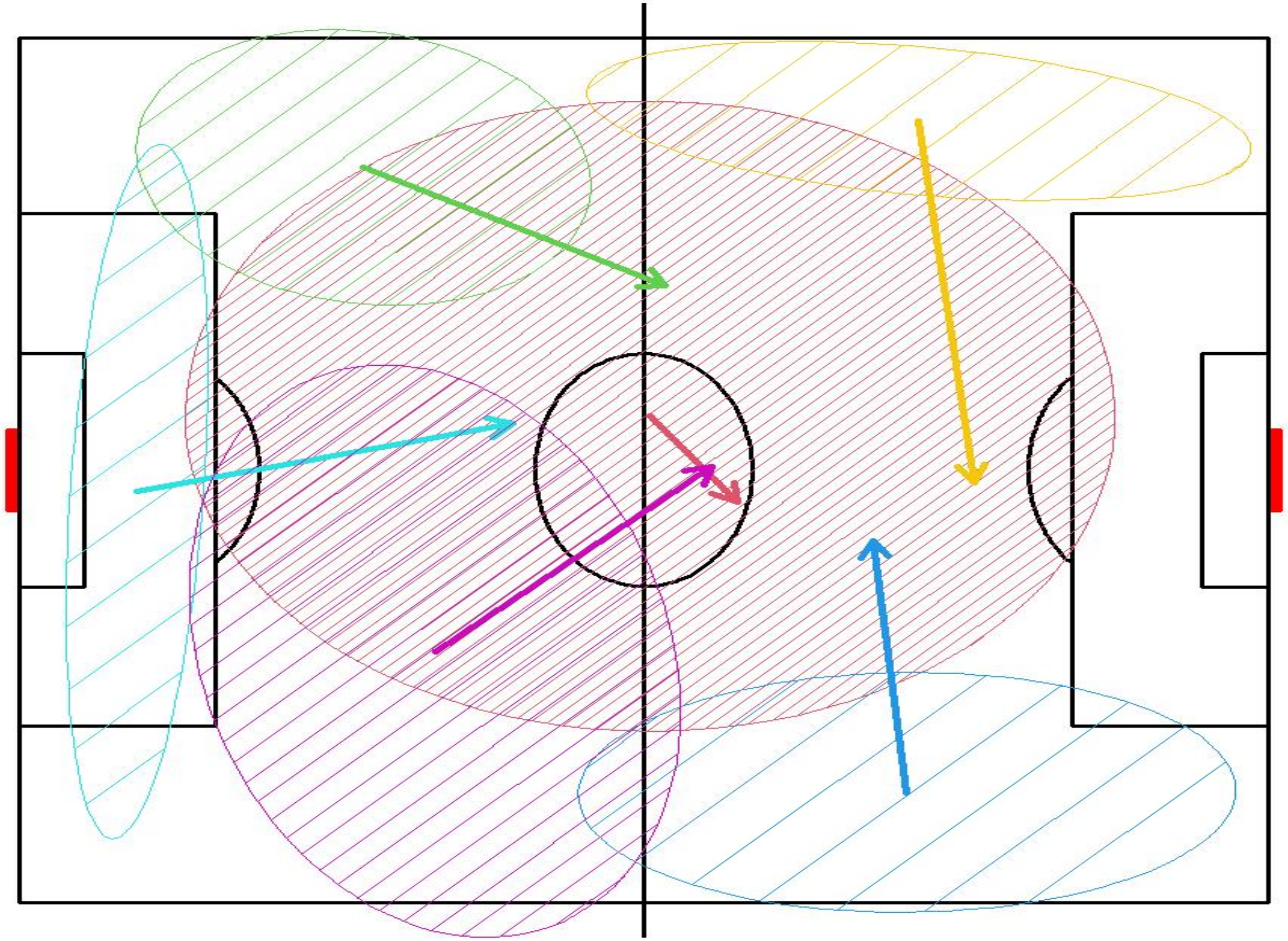
Ellipse:
90% region of the pass position

Direction of arrow:
mean direction of the pass in the position

Length of arrow:
mean resultant length of the pass in the position

Thickness of color:
proportional to the mixing parameter

Result



Ellipse:
  90% region of the pass position

Direction of arrow:
  mean direction of the pass in the position

Length of arrow:
  mean resultant length of the pass in the position

Thickness of color:
  proportional to the mixing parameter

Concluding remarks

Proposed methods have merits and demerits;

- □The marginals can be specified without complicated functions, and the conditionals belong to a known distribution family.
- □For special case, the moment and correlation are expressed in closed-forms.
- □Inference based on likelihood is easy and computational cost is not large.
- ■For the proposed construction, the linear part must be symmetric.

Concluding remarks

Proposed methods have merits and demerits;

- □The marginals can be specified without complicated functions, and the conditionals belong to a known distribution family.
- □For special case, the moment and correlation are expressed in closed-forms.
- □Inference based on likelihood is easy and computational cost is not large.
- ■For the proposed construction, the linear part must be symmetric.

However, for the transformed distribution

$$f(y,\theta) = 2G\left(\lambda\, w\left(\frac{2F_Y(x)}{\tau}, \theta - \mu_\Theta\right)\right) f_Y(y) f_\Theta(\theta - \mu_\Theta)$$

the marginals are $f_Y(y)$ and $f_\Theta(\theta - \mu_\Theta)$.

Concluding remarks

$$f(x,\theta) = 2G\left(\lambda\, w\left(\frac{x}{\tau}, \theta - \mu_\Theta\right)\right) f_X(x) f_\Theta(\theta - \mu_\Theta)$$

⇩ $f_X(x) = \frac{1}{2}$ uniform distribution on $[-1,1]$

$$f(x,\theta) = G\left(\lambda\, w\left(\frac{x}{\tau}, \theta - \mu_\Theta\right)\right) f_\Theta(\theta - \mu_\Theta)$$

⇩ $Y = F_Y^{-1}(X)$ for arbitrary distribution function $F_Y(\cdot)$

$$f(y,\theta) = 2G\left(\lambda\, w\left(\frac{2F_Y(x)}{\tau}, \theta - \mu_\Theta\right)\right) f_Y(y) f_\Theta(\theta - \mu_\Theta)$$

References

- T. Abe, A. Pewsey Sine-skewed circular distributions, Statistical Papers 52, 683-707, 2011.
- T. Imoto, T. Abe, Simple construction of a toroidal distribution from independent circular distributions, Journal of Multivariate Analysis, 2021.
- S. R. Jammalamadaka and Y. R. Sarma. A correlation coefficient for angular variables. In K. Matusita, editor, Statistical Theory and Data Analysis II, pages 349--364. Amsterdam: Elsevier, 1988.
- R. A. Johnson, T. E. Wehrly, Measures and models for angular correlation and angular-linear correlation, Journal of the Royal Statistical Society: Series B 39, 222-229, 1977.
- M. C. Jones, A. Pewsey, S. Kato, On a class of circulas: copulas for circular distributions, Annals of the Institude of Statistical Mathematics 67, 843-862, 2015.
- K. V. Mardia, Linear-circular correlation coefficients and rhythmometry, Biometrika 63, 403-405, 1976.
- L. Pappalardo, P. Cinitia, A. Rossi, E. Massucco, P. Ferragina, D. Pedreschi, F. Giannotti, A public data set of spatio-temporal match events in soccer competitions, Nature Scientific Data 6:236, 2019a.
- L. Pappalardo, P. Cinitia, P. Ferragina,E. Massucco,D. Pedreschi, F. Giannotti, PlayerRank: Data-driven performance evaluation and player ranking in soccer via a machine learning approach. ACM Transactions on Intelligent Systems and Technologies 10, 5, Article 59, 2019b.
- T. E. Wehrly, R. A. Johnson, Bivariate models for dependence of angular observations and related Markov process, Biometrika 66, 255-256, 1979.

Thank you for your attention !