Probability Distributions on the Torus with Applications to Bioinformatics

Jochen Voss

University of Leeds

8th April 2010

joint work with Kanti V. Mardia

Outline

Motivation

There are many problems in applied statistics where it is beneficial to use probabilistic methods for angles. This is part of **directional statistics**.

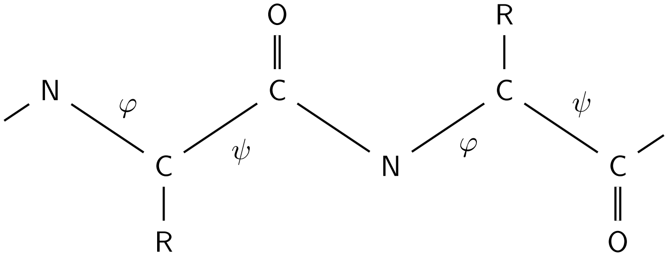
- distribution of a single angle = distribution on a circle
- joint distribution of several angles = distribution on a torus $\mathbb{T}^d$

Application areas include

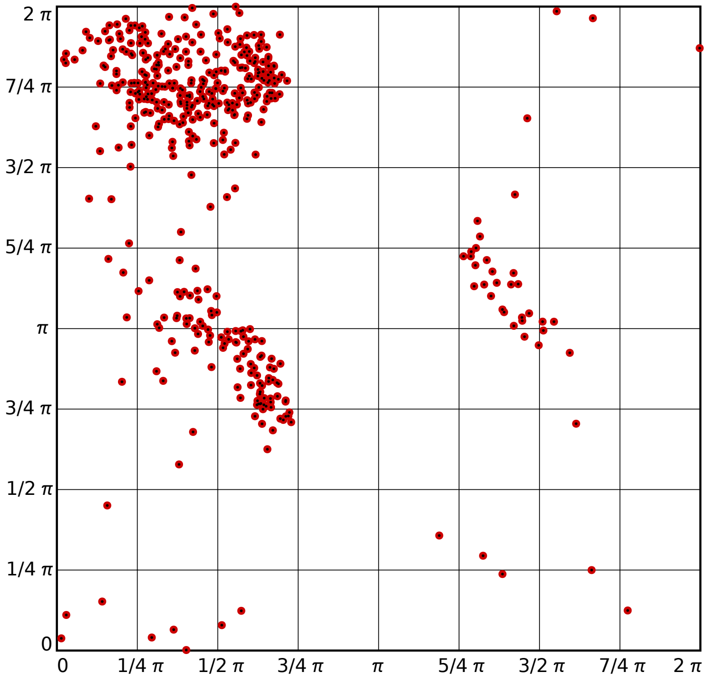
- astronomy
- geology
- biochemisty

In this talk I'll briefly consider some examples from biochemisty, and in particular distributions of angles describing the conformation of proteins or RNA molecules.

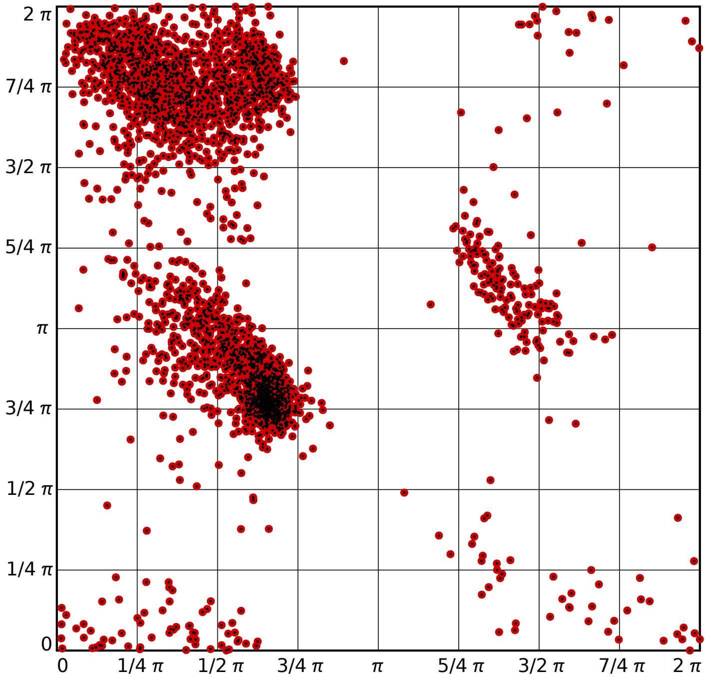
Example: backbone angles of proteins

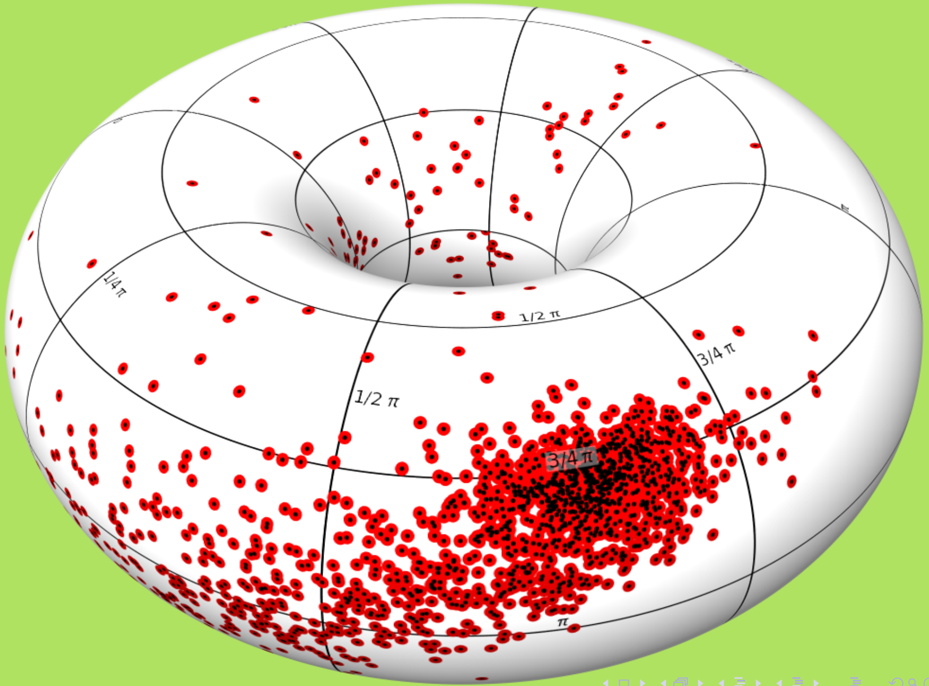


- Most of the geometric structure of the protein backbone is described by the sequence of dihedral angles φ_1, ψ_1, φ_2, ψ_2, ...
- By creating a scatter-plot of (φ, ψ) (Ramachandran plot) for naturally occuring proteins, one can see that (φ, ψ) have a non-trivial joint distribution on the torus $\mathbb{T}^2$.
- This forms the basis for probabilistic models of protein structure.



2 π
7/4 π
3/2 π
5/4 π
π
3/4 π
1/2 π
1/4 π
0
0
1/4 π
1/2 π
3/4 π
π
5/4 π
3/2 π
7/4 π
2 π


2 π
7/4 π
3/2 π
5/4 π
π
3/4 π
1/2 π
1/4 π
0
0
1/4 π
1/2 π
3/4 π
π
5/4 π
3/2 π
7/4 π
2 π


1/4 π
1/2 π
1/2 π
3/4 π
3/4 π
π

Angular Distributions

Commonly used distributions of one angle include:

- uniform: $\theta \sim \mathcal{U}[0, 2\pi]$
- von Mises distribution: density $\varphi(\theta) = \frac{1}{Z(\kappa)} \exp\big(\kappa \cos(\theta - \mu)\big)$
- Bingham distribution
- wrapped normal distribution:
  $\theta = X \mod 2\pi$ where $X \sim \mathcal{N}(\mu, \sigma^2)$
- wrapped Cauchy distribution:
  $\theta = X \mod 2\pi$ where X is Cauchy distributed on $\mathbb{R}$

Differences to probability distributions on $\mathbb{R}$:

- The usual definitions of "expectation", "variance" *etc.* fail.
  (What is the mean of a uniform distributions on angles?)
- There are no "tails".

Commonly used distributions of several angles include:

- the "full" bivariate von Mises distribution
- the multi-variate von Mises distribution (see next slide)
- wrapped multivariate normal distribution $\mathcal{N}(\mu, \Sigma)$

Such distributions can be used as building blocks to model distributions like in the Ramachandran plot (*e.g.* as a mixture of several von Mises distributions).

Remark. These are distributions on the torus. Distributions on the sphere are different!

The Sine Model

The *multivariate von Mises distribution* on the d-dimensional torus $\mathbb{T}^d$ has density

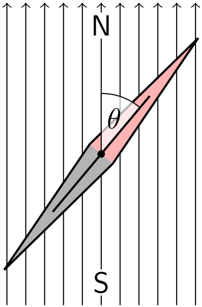
$$\varphi(\theta;\mu,\kappa,\Lambda)=\frac{1}{Z(\kappa,\Lambda)}\exp\bigl(\kappa^\top c(\theta)+\frac{1}{2}s(\theta)^\top\Lambda\, s(\theta)\bigr)$$

where

- $c_i(\theta)=\cos(\theta_i-\mu_i)$ and $s_i(\theta)=\sin(\theta_i-\mu_i)$ for $i=1,\dots,d$,
- $\mu\in\mathbb{T}^d$ is the location parameter,
- $\kappa\in\mathbb{R}_+^d$ determines "concentration",
- $\Lambda=(\lambda_{ij})\in\mathbb{R}^{d\times d}$ with $\Lambda^\top=\Lambda$ and $\lambda_{ii}=0$ for $i=1,\dots,d$, and
- $Z(\kappa,\Lambda)$ is the normalisation constant.

Remark. For $d=1$ one has $Z(\kappa,\Lambda)=2\pi I_0(\kappa)$ where I_0 is the modified Bessel function of order 0. For $d>1$ there is no closed-form expression for the normalising constant Z.

Example. The motion of a compass needle in a magnetic field under the influence of white noise is described by the following stochastic differential equation:



$$d\theta(t) = -\alpha \sin\big(\theta(t)\big)\, dt + \sigma\, dB(t)$$

where

- B: a Brownian motion
- $\alpha > 0$: strength of the magnetic field
- $\sigma > 0$: strength of noise

It is easy to check that the equilibrium distribution of $\theta(t)$ is a one-dimensional von Mises distribution with $\kappa = 2\alpha/\sigma^2$ and $\mu = 0$.

Normal Approximation. By Taylor-approximation we have $\cos(\theta) \approx 1 - \frac{1}{2}\theta^2$ and $\sin(\theta) \approx \theta$. Thus

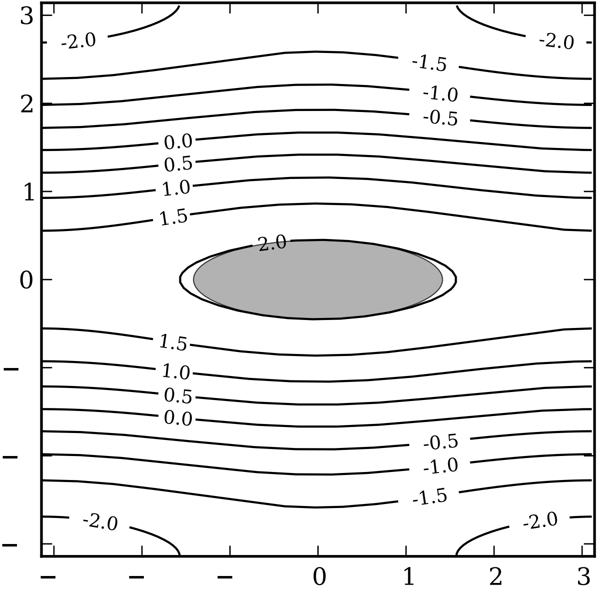
$$\begin{aligned}\kappa^\top c(\theta) + \frac{1}{2} s(\theta)^\top \Lambda\, s(\theta) &\approx \sum_{i=1}^{d} \kappa_i (1 - \frac{1}{2}\theta_i^2) + \frac{1}{2} \sum_{i,j=1}^{d} \theta_i \lambda_{ij} \theta_j \\ &= -\frac{1}{2}\theta^\top \Sigma^{-1}\theta + \text{const.}\end{aligned}$$

where

$$\Sigma^{-1} = \operatorname{diag}(\kappa) - \Lambda.$$

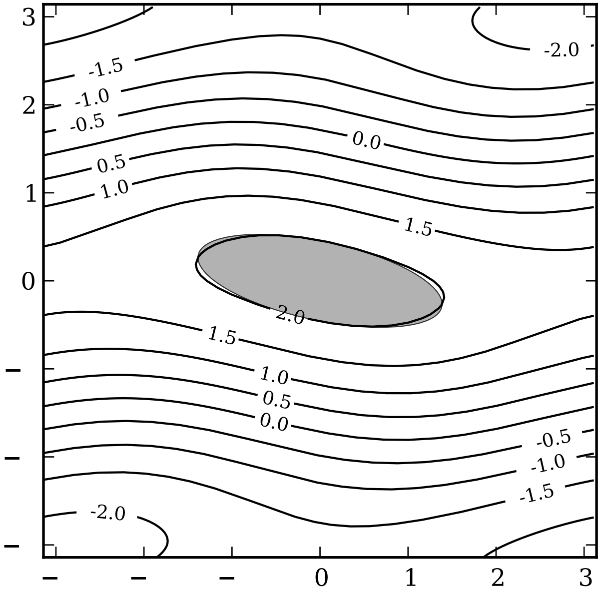
Result: For "big" κ, the multivariate von Mises distribution "converges" to the wrapped $\mathcal{N}(\mu, \Sigma)$ distribution.

Lemma. Assume that the matrix Σ^{-1} is positive definite. Then the global maximum of the von Mises density φ is attained at $\theta = \mu$ and φ has no other (local) maxima.



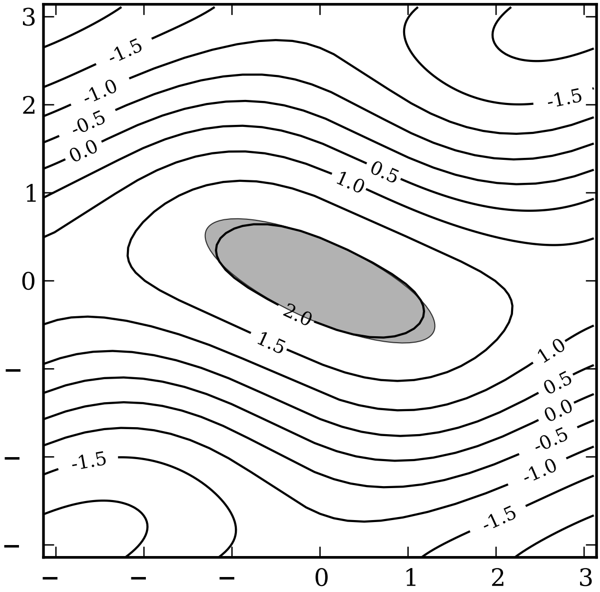
$$\kappa = \begin{pmatrix} 0.200 \\ 2.000 \end{pmatrix}$$

$$\Lambda = \begin{pmatrix} 0 & -0.000 \\ -0.000 & 0 \end{pmatrix}$$



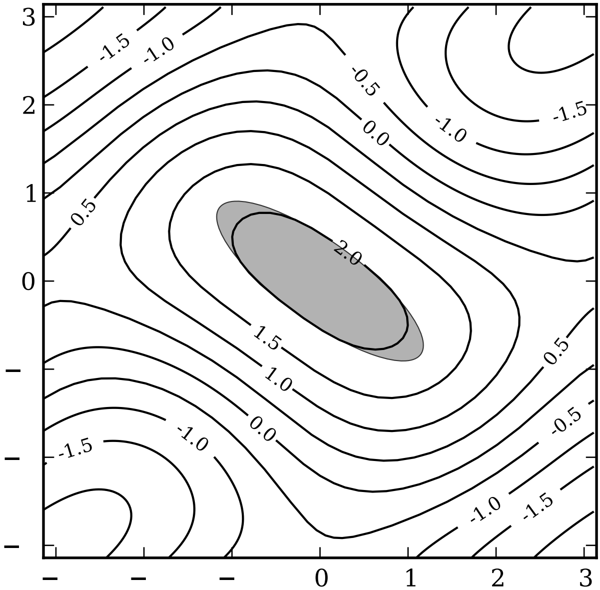
$$\kappa = \begin{pmatrix} 0.278 \\ 1.922 \end{pmatrix}$$

$$\Lambda = \begin{pmatrix} 0 & -0.366 \\ -0.366 & 0 \end{pmatrix}$$



$$\kappa = \begin{pmatrix} 0.498 \\ 1.702 \end{pmatrix}$$

$$\Lambda = \begin{pmatrix} 0 & -0.669 \\ -0.669 & 0 \end{pmatrix}$$



$$\kappa = \begin{pmatrix} 0.822 \\ 1.378 \end{pmatrix}$$

$$\Lambda = \begin{pmatrix} 0 & -0.856 \\ -0.856 & 0 \end{pmatrix}$$

Small/vanishing concentration. If κ is "small", the von Mises distribution can have a surprising number of modes. Example: for

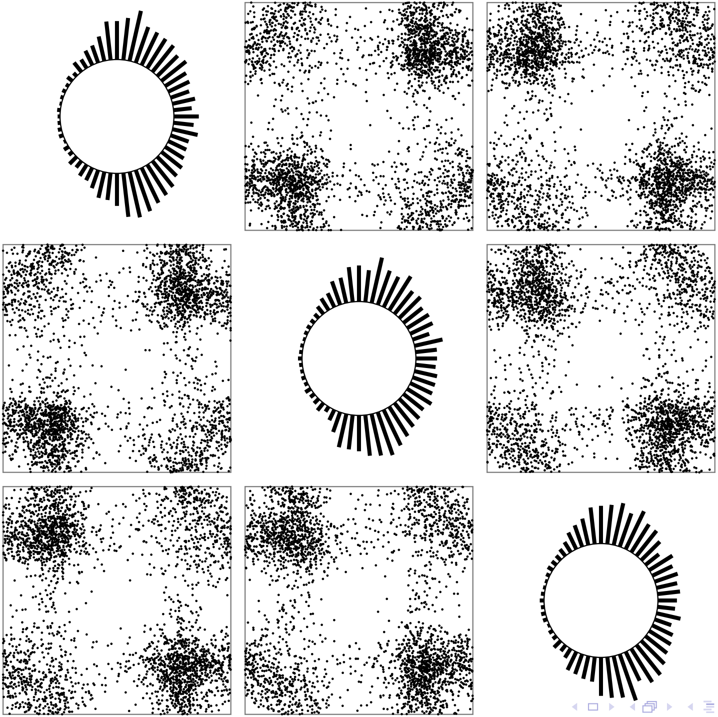
$$\kappa = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \qquad \Lambda = \begin{pmatrix} 0 & -5 & 5 \\ -5 & 0 & 5 \\ 5 & 5 & 0 \end{pmatrix}$$

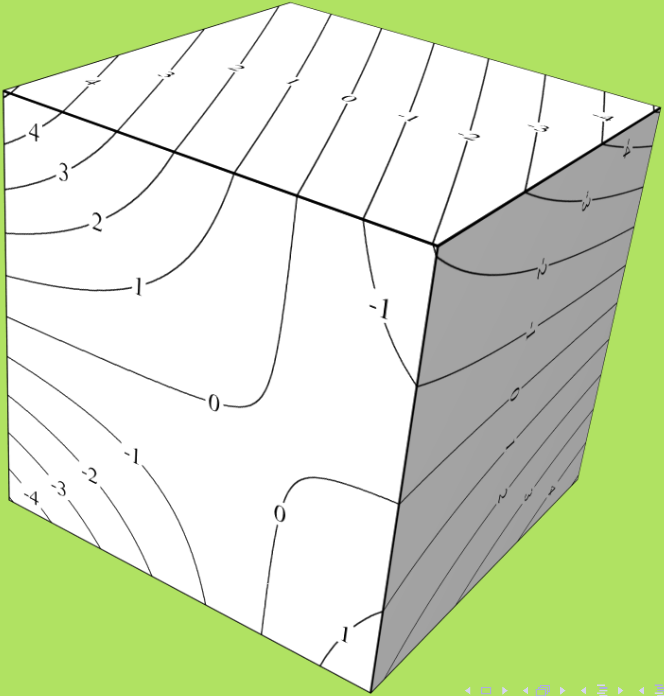
the distribution has 6 isolated modes. These cases will **not** be useful for modelling!

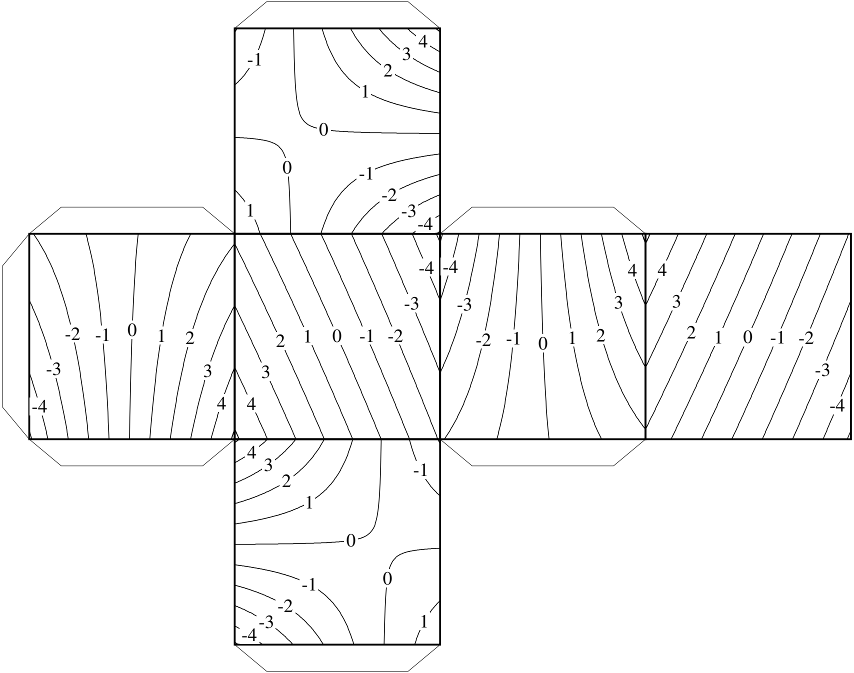
In the extreme case $\kappa = 0$ the density of the von Mises distribution simplifies to

$$\varphi(\theta; \mu, 0, \Lambda) = \frac{1}{Z(0, \Lambda)} \exp\left(\frac{1}{2} s^\top \Lambda\, s\right) =: g(s)$$

where $s = \big(\sin(\theta_1), \ldots, \sin(\theta_d)\big) \in [-1, 1]^d$. It transpires that at the maxima of the density, s is located in the corners of the cube $[-1, 1]^d$.




4
3
2
1
0
-1
-1
-2
-3
-4
0
1


-1 4 3 2 1 0 0 -1 -2 1 -3 -4
-2 -1 0 1 2 -3 3 -4 4
-4 -3 2 1 0 -1 -2 3 4
-4 4 -3 3 -2 -1 0 1 2
4 3 2 1 0 -1 -2 -3 -4
4 3 2 -1 1 0 -1 0 -2 -3 1 -4

Outlook

For “large” κ the multivariate von Mises distribution can be used as a building block for modelling distributions on the torus. Further questions include:

- How to sample from a multivariate von Mises distribution?
- How to perform statistical inference, *e.g.* fitting of parameters?

References

- Kanti V. Mardia and Peter E. Jupp, *Directional statistics*.
  Wiley Series in Probability and Statistics, 2000
- Wouter Boomsma, Kanti V. Mardia, Charles C. Taylor, Jesper Ferkinghoff-Borg, Anders Krogh, Thomas Hamelryck, *A generative, probabilistic model of local protein structure*.
  PNAS, vol. 105 (26), pp. 1-6, 2008
- Kanti V. Mardia, Gareth Hughes, Charles C. Taylor, Harshinder Singh, *A multivariate von Mises distribution with applications to bioinformatics*.
  The Canadian Journal of Statistics, vol. 36 (1), pp. 99-109, 2008