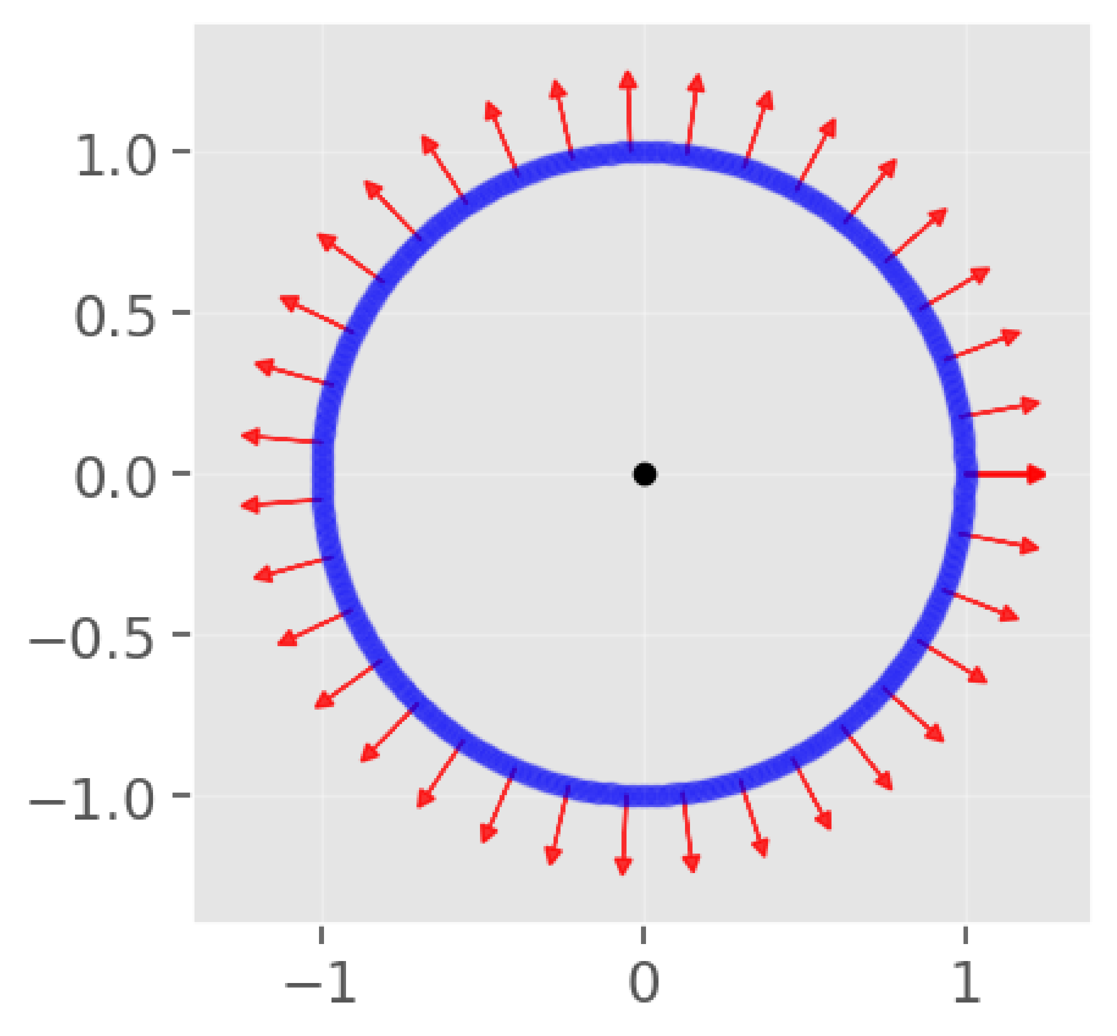
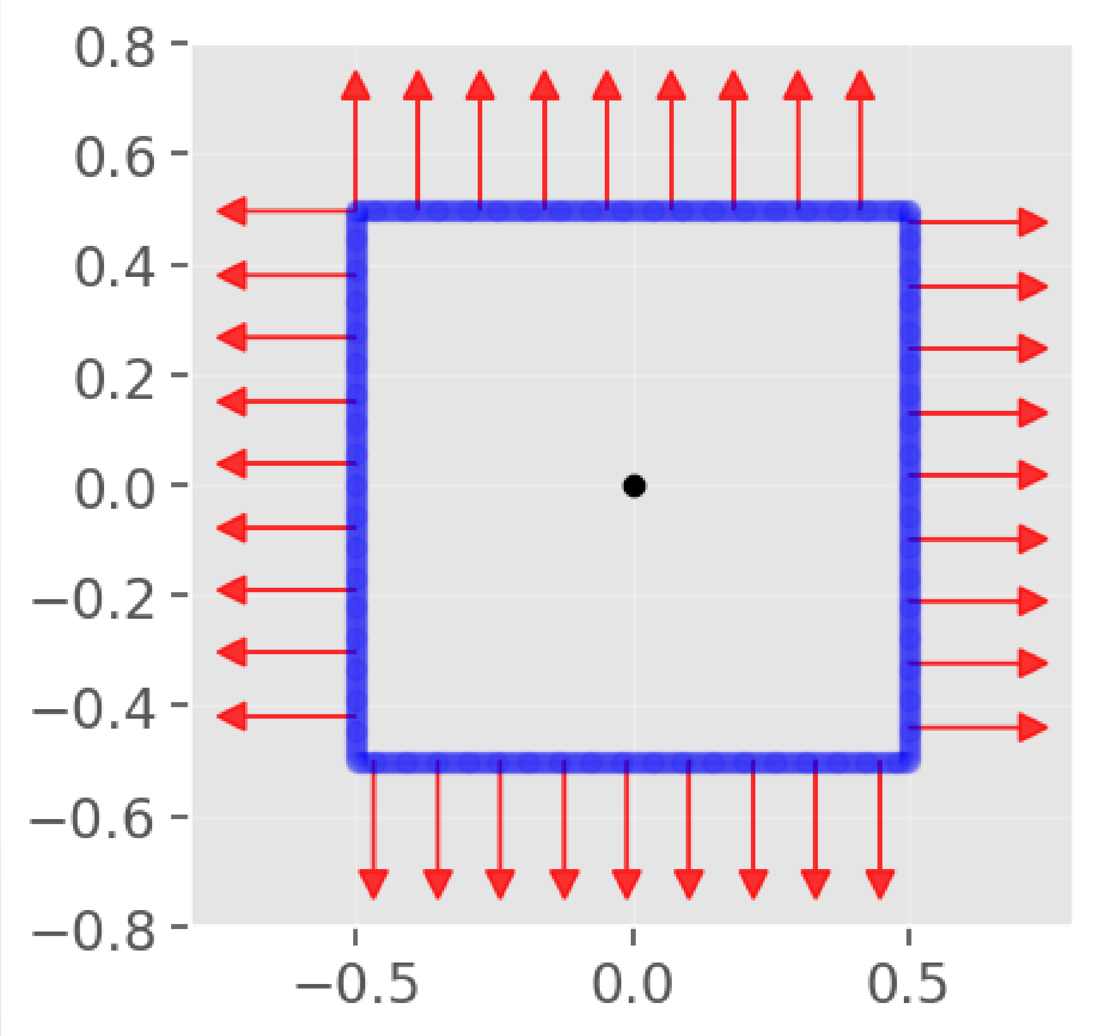
フォン=ミーゼス分布が解き明かす図形の特徴

～データサイエンスの幾何学への応用～





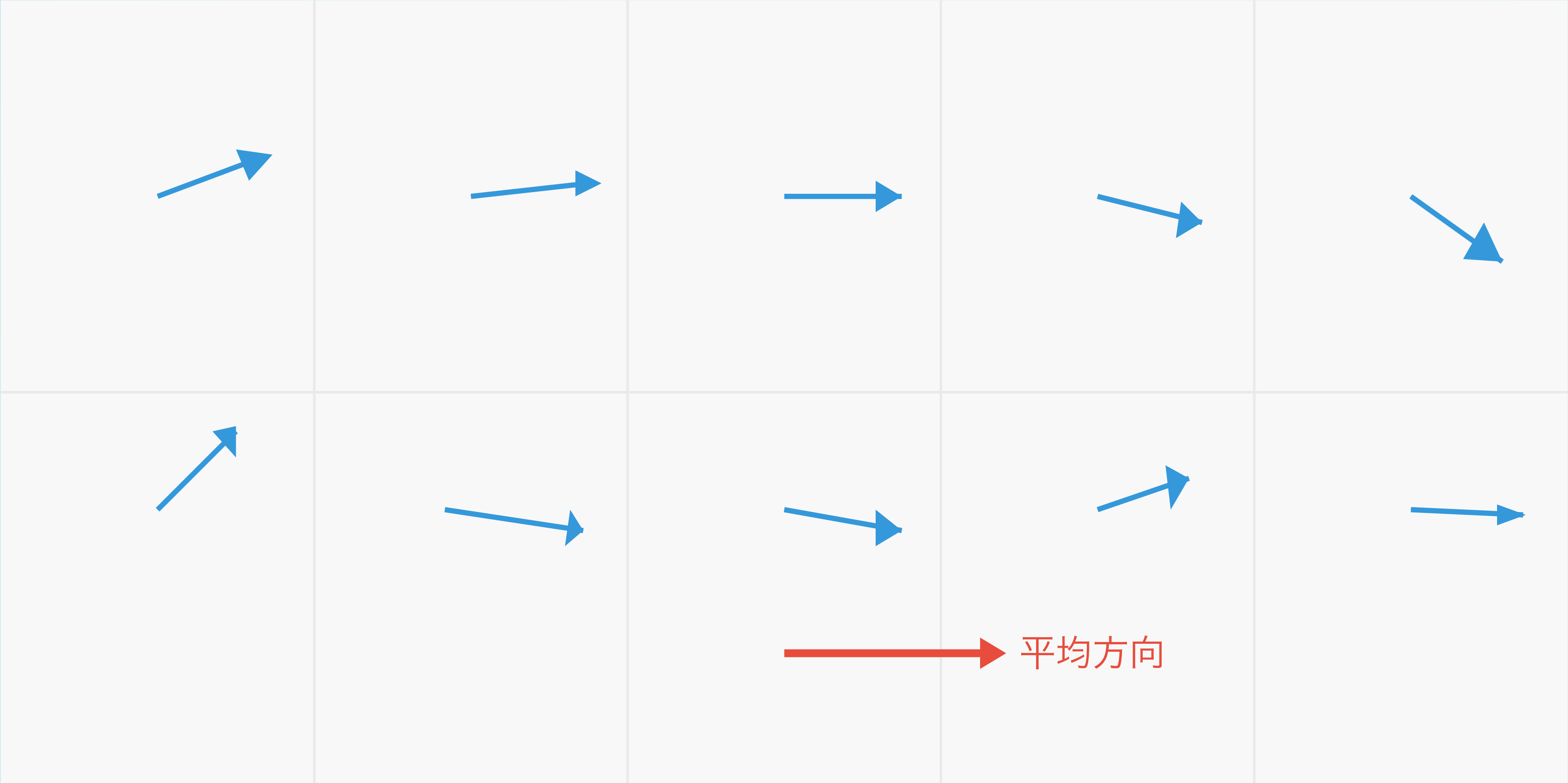
自己紹介

- さめ(meg-ssk)
- 👨‍💻 フリーランスのソフトウェアエンジニア
- 得意分野:
  - 📸 コンピュータビジョン(画像認識/点群処理)
  - 🌍 空間情報処理 (地理情報/リモートセンシング)
  - ☁️ クラウドインフラ設計/IaC (AWS, GCP)
- GitHub
- YouTube
- Speaker Deck

ハイライト

- フォン=ミーゼス分布は「方向の分布」を表す分布関数である
- 重要な応用例を数多く持つ
  - 大きさだけではなく方向を持つ量(ベクトル)の分析
    - 風向や図形の法線分布など
- **今日は基礎的なコンセプトとオープンデータを使った分析例を示します！**
- 方向統計学(Directional Statistics)の基礎を紹介します

簡単な例



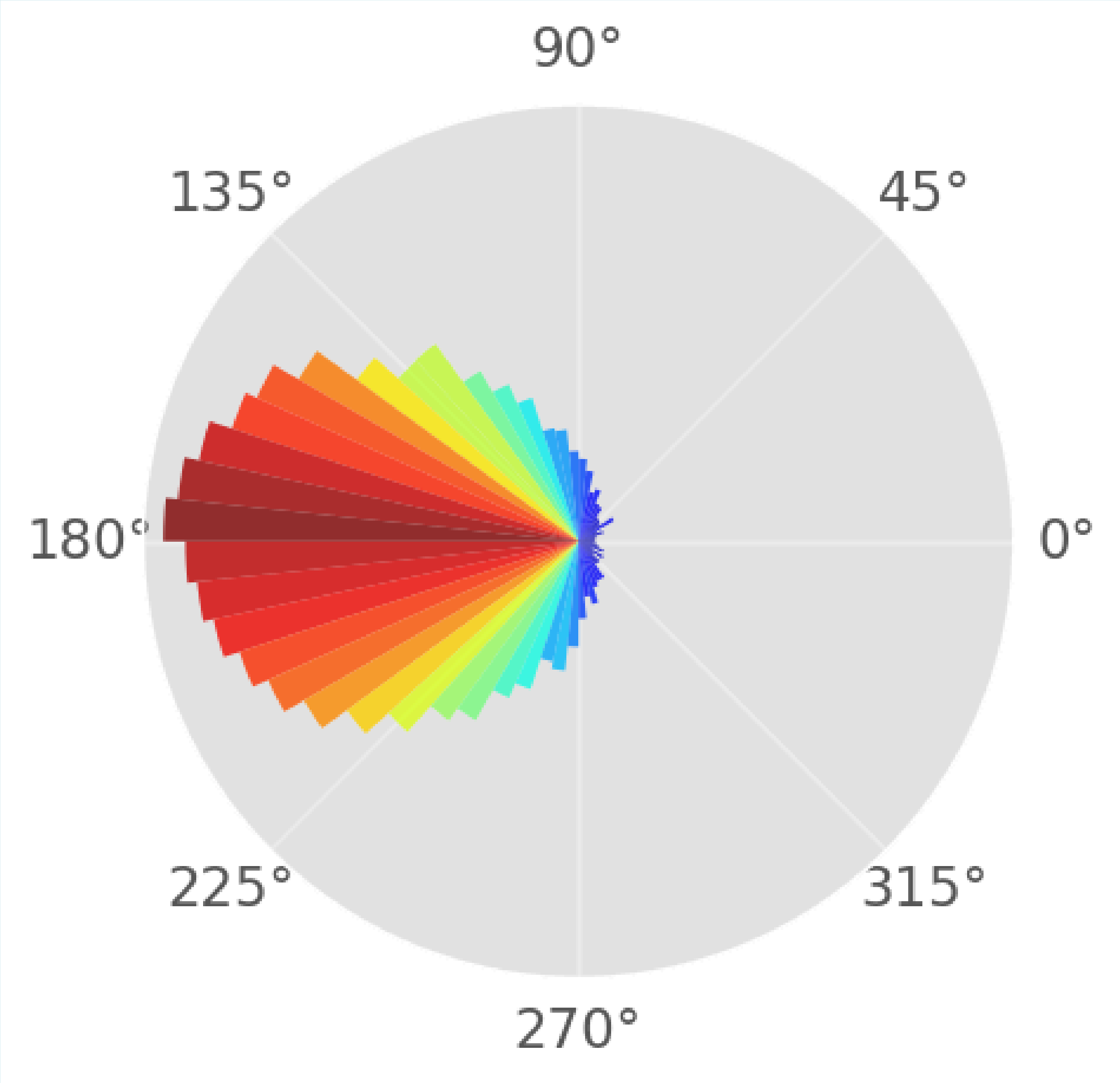
- 平均すれば右を向いているが、ひとつひとつのサンプルの向きは揺らぐ
- 方向の分布を表す分布関数とそれを特徴付けるパラメータは？

フォン=ミーゼス分布

$$f(\theta) = \frac{\exp(\kappa \cos(\theta - \mu))}{2\pi I_0(\kappa)}$$

- 言うなれば2次元のベクトルの向きの正規分布
  - μ: 平均
  - κ: 集中度
  - $I_0(.)$: 第1種ベッセル関数

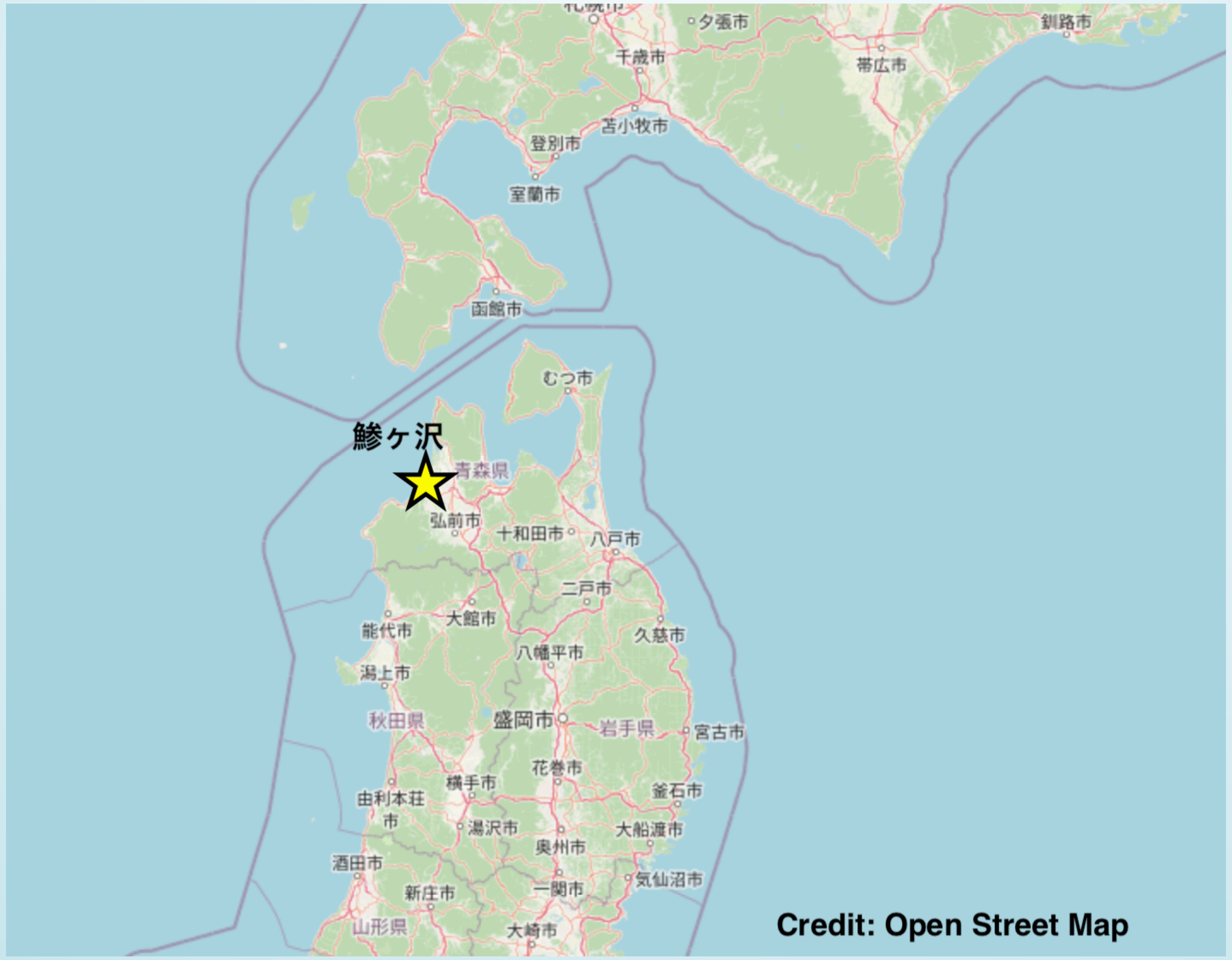
フォン=ミーゼス分布の可視化



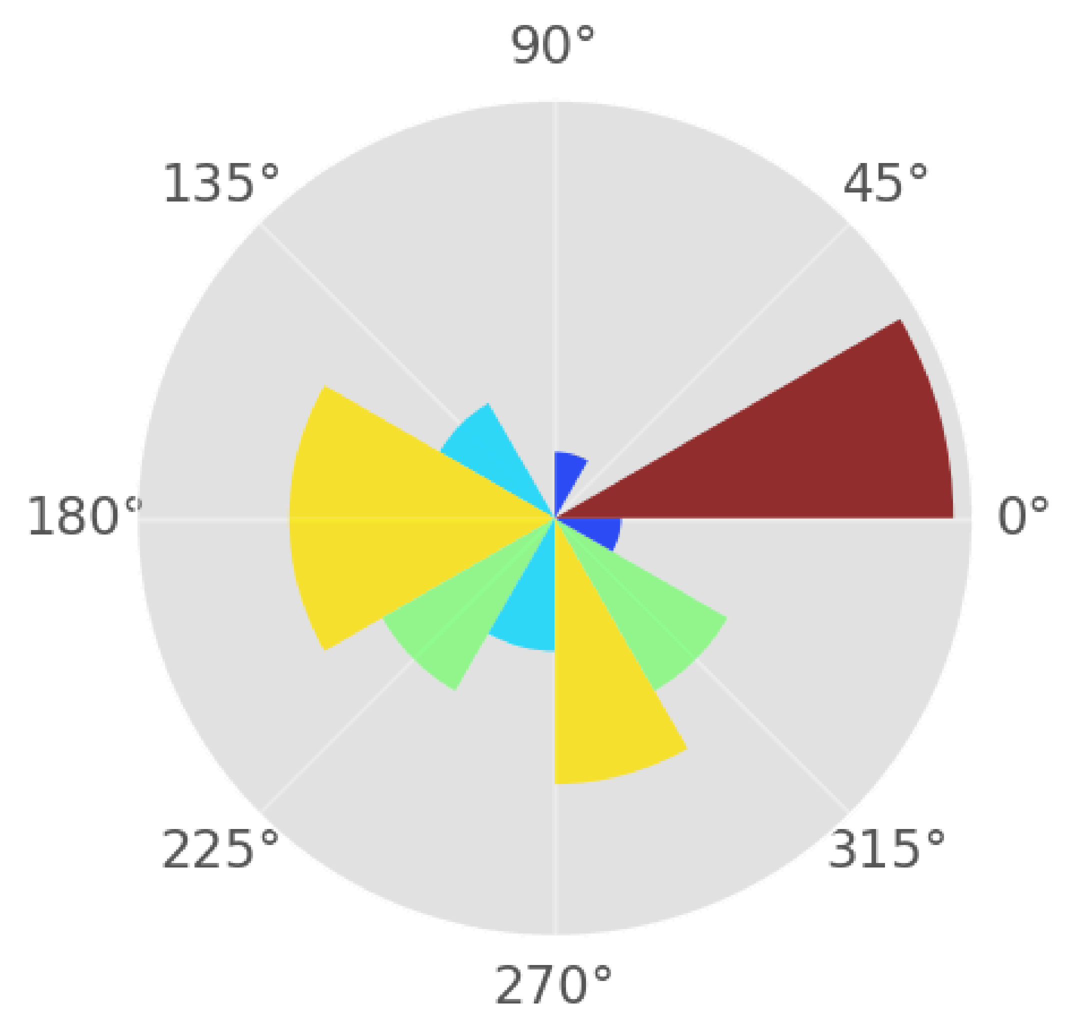
- $\mu = 180°, \kappa = 1.5$ の例を極座標ヒストグラムで表現

実用例: 風向の分布解析

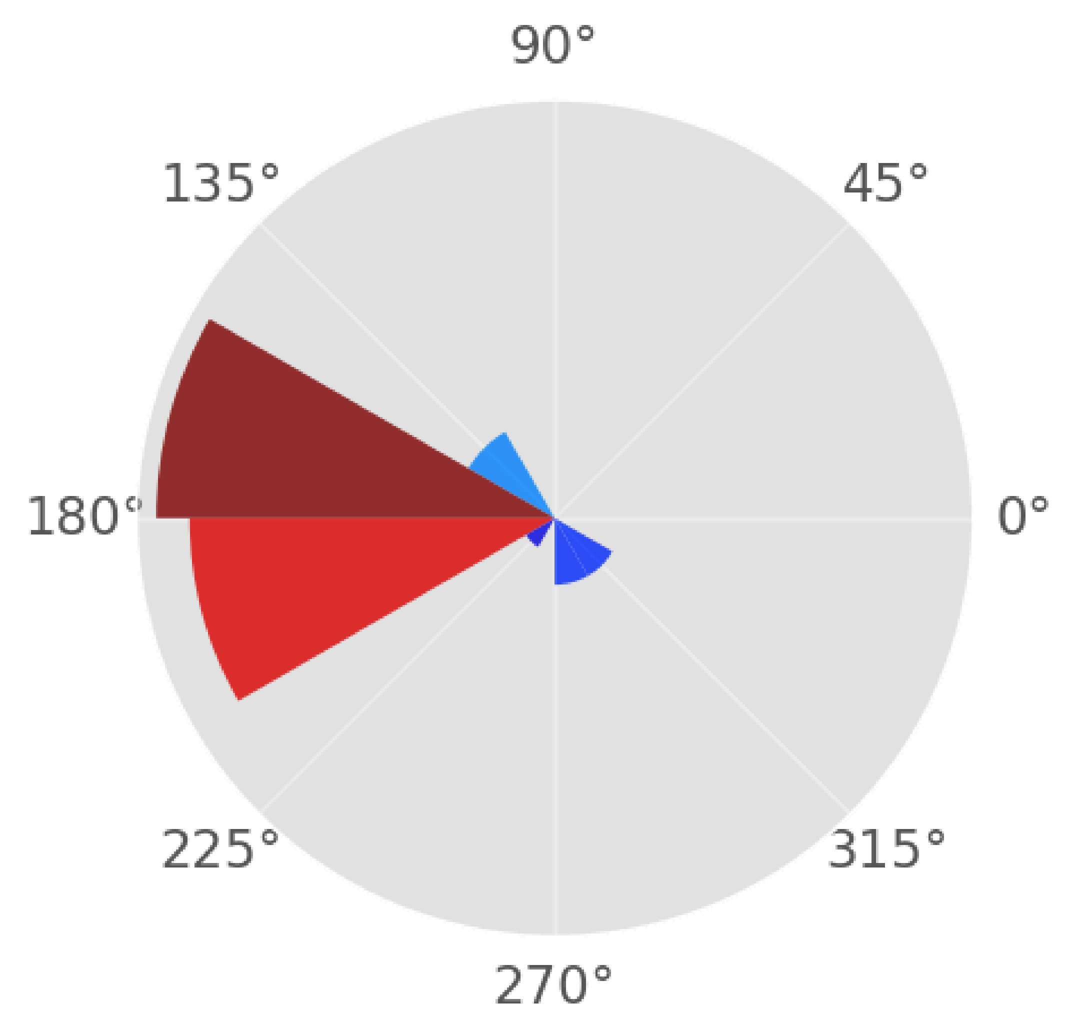
- 日本海側は冬に北西から季節風が吹く
  - 気象庁が公開している青森県鰺ヶ沢町の風向分布のデータを利用して検証



2024年6月と12月の風向分布



6月の日別最頻風向分布



12月の日別最頻風向分布

- 西方向から風が吹く頻度が12月の方が高い
  - 本来ならより詳細な検討が必要だが割愛

風向のパラメタライズ

- 風向がフォン=ミーゼス分布に従うと仮定し、6月と12月の風向の平均と集中度を推定

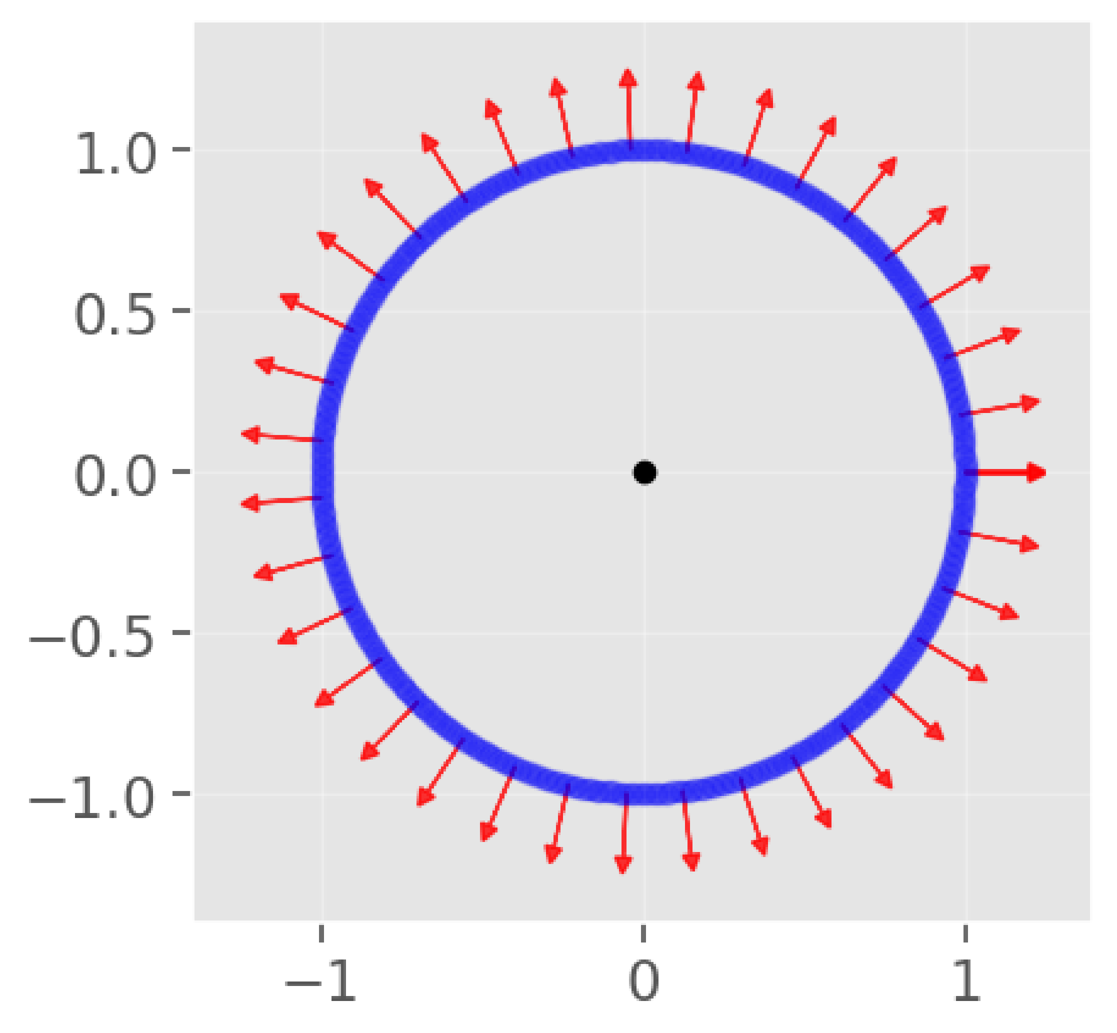
- 6月
  - 平均: $\mu \simeq 242^\circ$
  - 集中度: $\kappa \simeq 0.46$

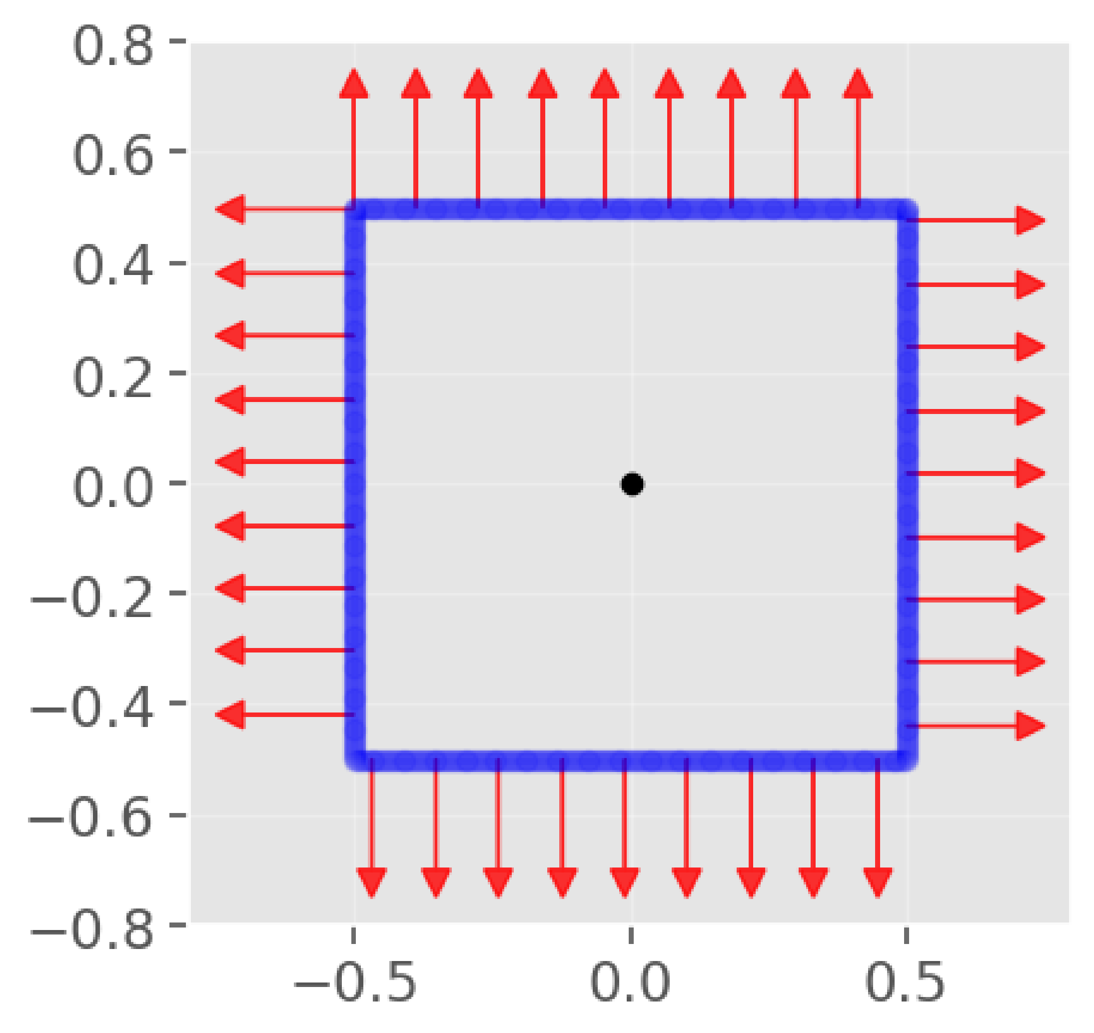
- 12月
  - 平均: $\mu \simeq 170^\circ$
  - 集中度: $\kappa \simeq 2.40$

- 6月は集中度が低く各方向に分散
- 12月は集中度が高く、西方向からの風の頻度が高い

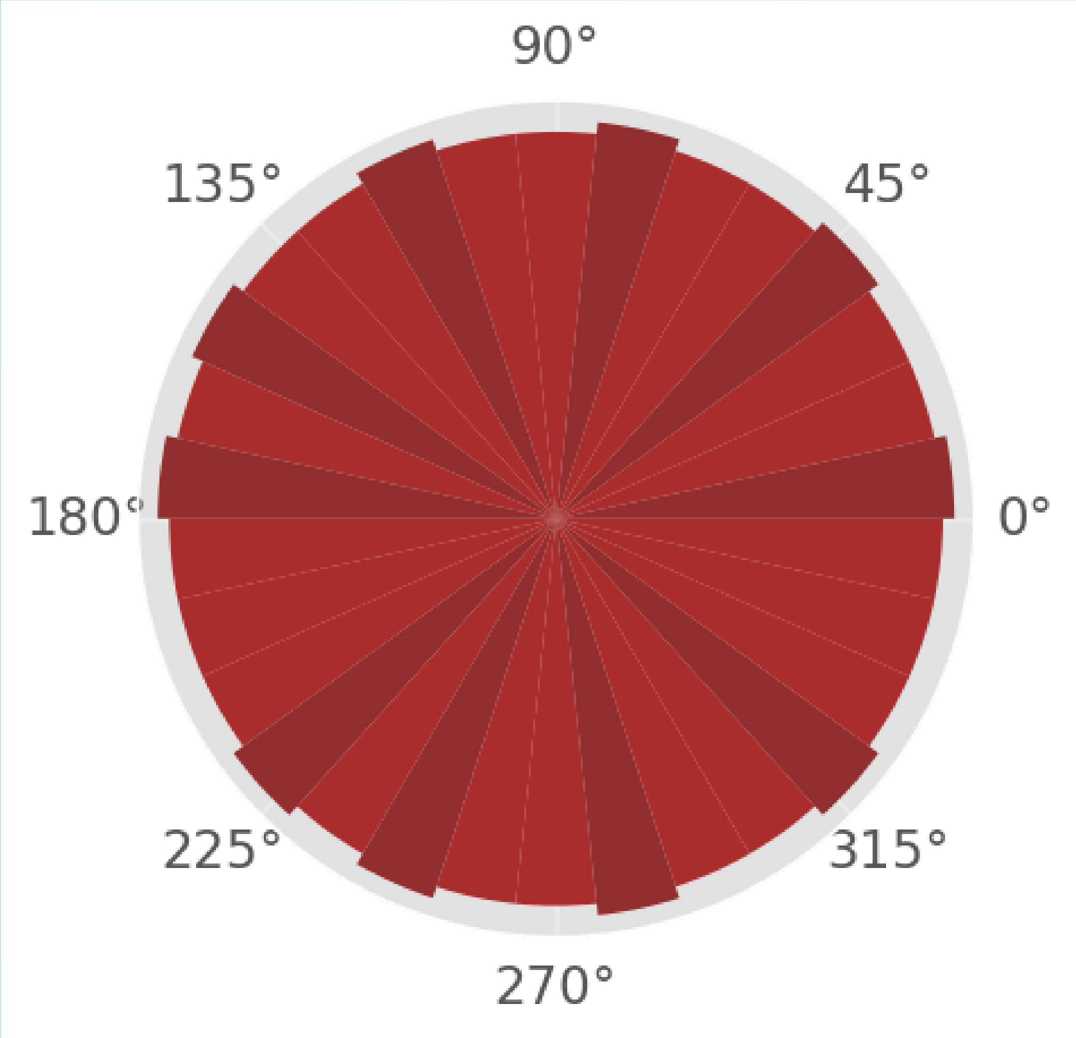
図形の法線方向の分布

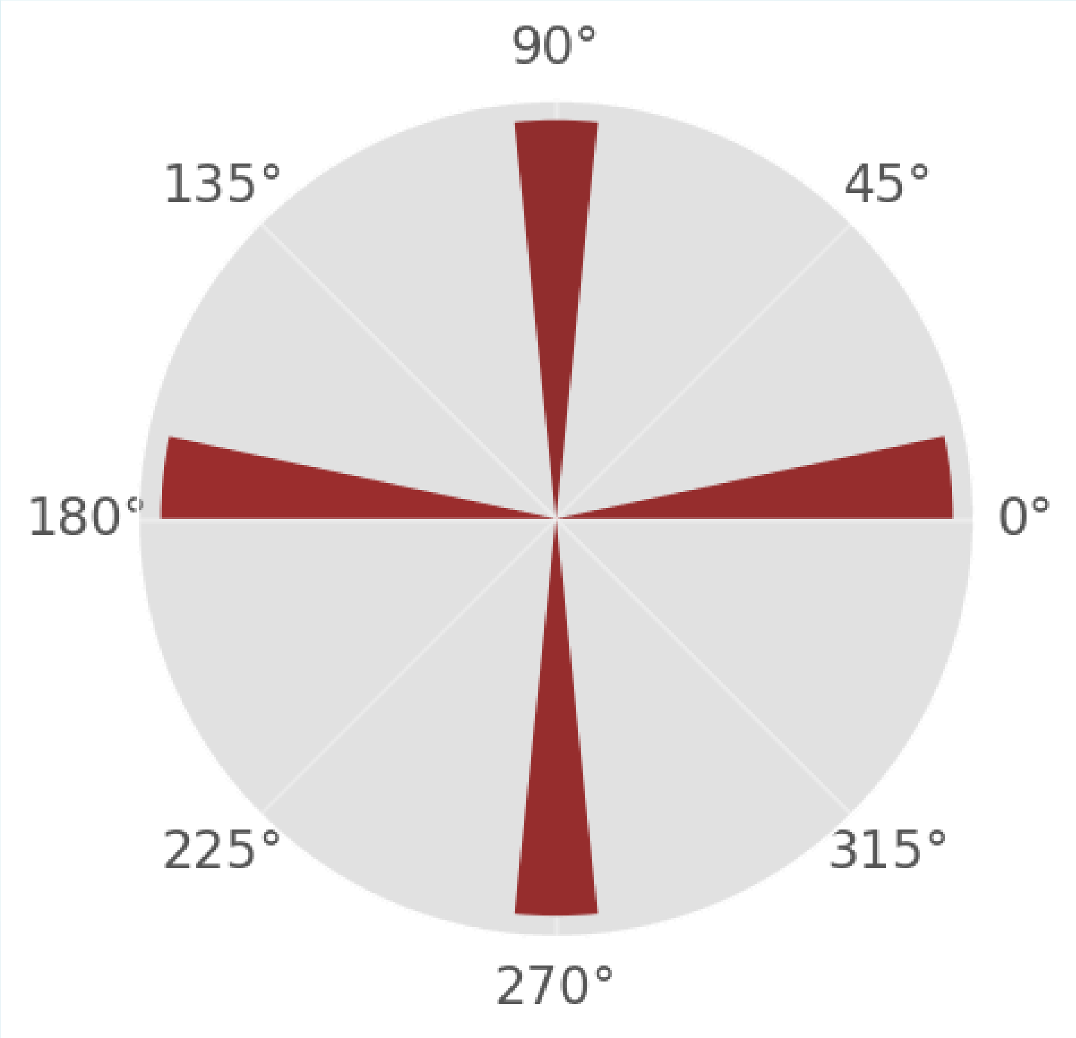
- 円と正方形の法線の分布を比較





円と正方形の法線分布の比較





- 円の法線分布は一様、集中度κは0
- 正方形の法線は4つピークを持つ。平均μは0度
- **法線分布から図形の情報を抽出できる！**

3次元への拡張: フォン=ミーゼス-フィッシャー分布

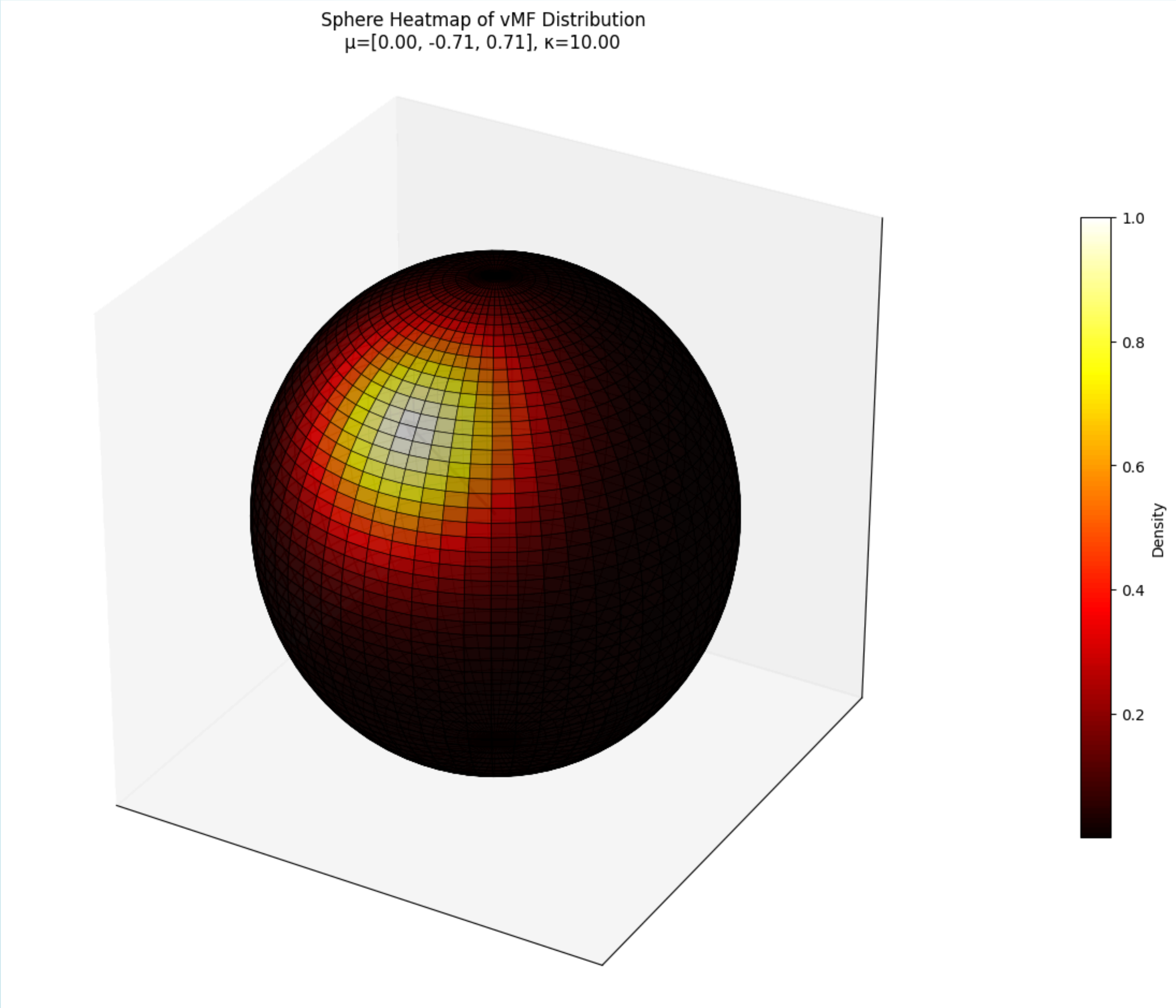
- フォン=ミーゼス分布は2次元平面での議論
- **フォン=ミーゼス-フィッシャー分布**は一般のd次元に拡張可能
  - ただし今日は3次元のみにフォーカス

$$f_3(\mathbf{x}) = \frac{\sqrt{\kappa}}{(2\pi)^{3/2} I_{1/2}(\kappa)} \exp(\kappa \langle \boldsymbol{\mu}, \mathbf{x} \rangle)$$

平均$\boldsymbol{\mu}$がベクトルになることに注意

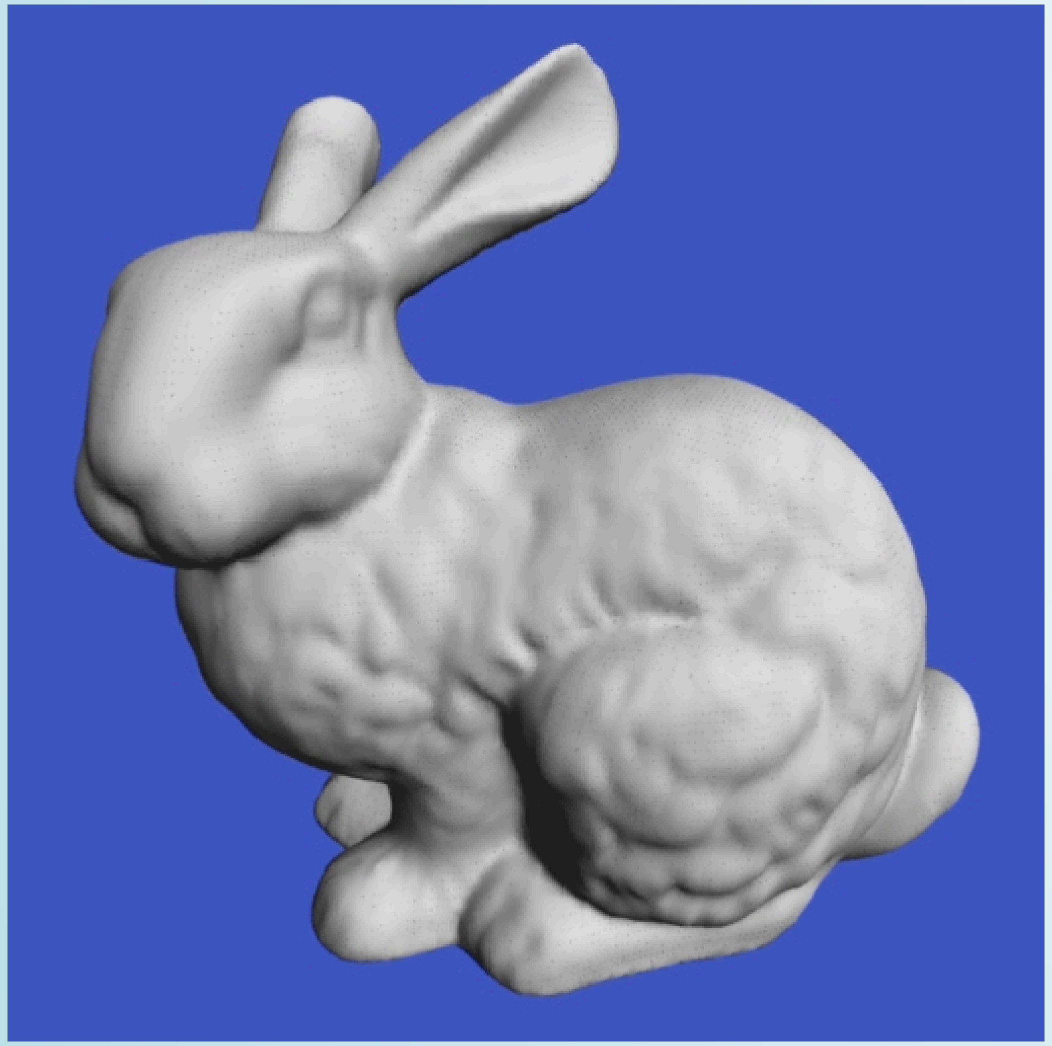
3次元のフォン＝ミーゼス-フィッシャー分布の可視化

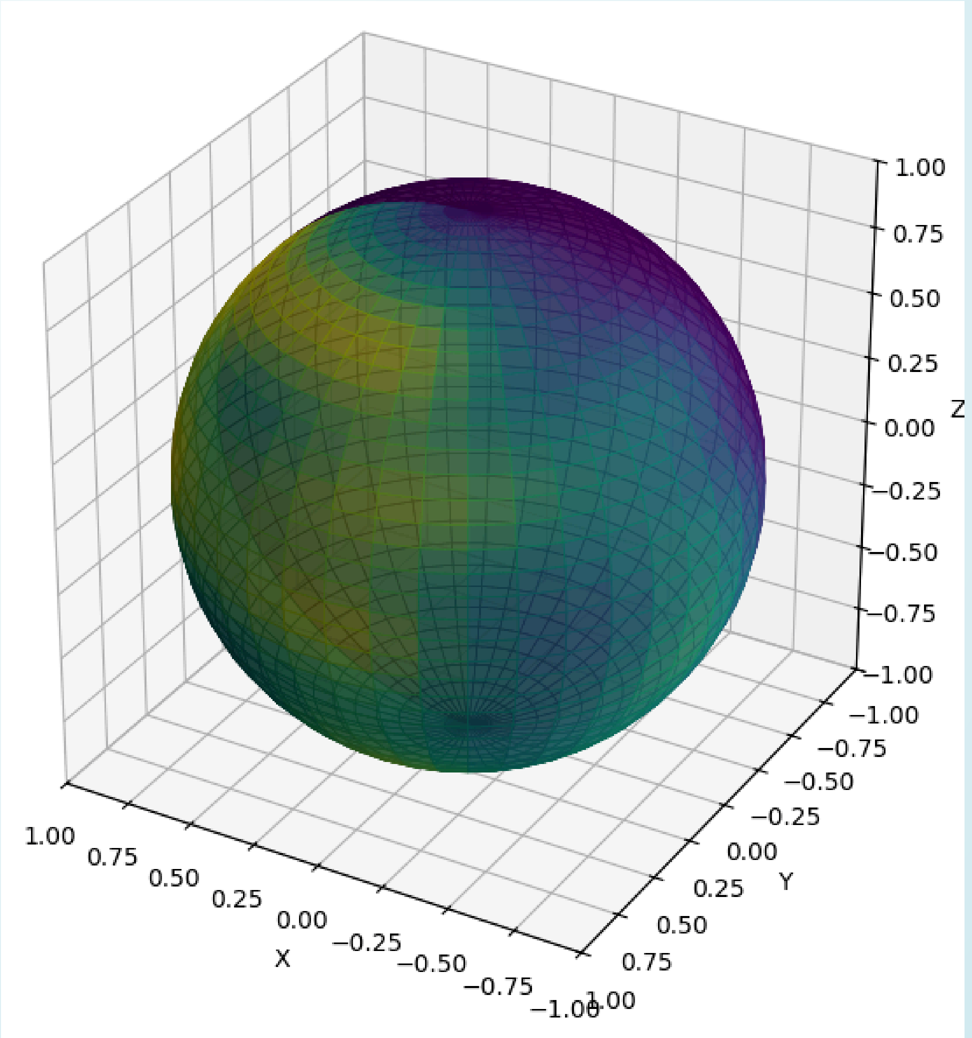
- 放線密度を球上のヒートマップで可視化



実用例: 点群データの法線分布

- スタンフォードバニーの法線分布
- どの方向にも分布する比較的均一な分布





- 有名なウサギの3Dモデル

どのような3Dモデルが特徴的な法線分布を持つか？

- 直方体に近い3Dモデルは6箇所にピークを持つ
  - 室内やビル群など
- 草原は上向きの法線(地表面)を多く持つが、方向の集中度は低い
  - 表面は荒いので法線の方向はあまり集中しない
  - 横向きの法線は少ない
- 球体はどの方向も均一に分布
  - バニーはほとんど球体 (「まず牛を球とします」は大体本当)

さらなる応用例

- フォン=ミーゼス-フィッシャー分布は本質的にはR^d空間のベクトルの分布を表す分布関数
- データサイエンスや機械学習モデルの開発では様々なベクトル化されたデータを扱う
- **フォン=ミーゼス-フィッシャー分布の平均方向と集中度を推定することで、ベクトル化されたデータの特徴を単純化して捉えることができる**
  - 角度の分布は位相の分布としても解釈可能 → 音声処理や信号処理の分野でも応用可能
  - 異常検知などに応用可能（......らしいです）

まとめ

- フォン＝ミーゼス-フィッシャー分布はベクトルの分布を表す分布関数である
- 平均方向と集中度の2つのパラメーターでベクトルの分布を特徴づけられる
- 今日は図形の特徴を定量化するための手法としての応用例を紹介した
- 幾何学的な応用に限らず、ベクトルの集合を分析する有力な手法である

補足

- 法線分布は図形の特徴を表現する方法のひとつ
- ガウス曲率などの幾何学的量も図形の特徴を表現するのに役立つ（3次元の場合）
- 複数の定量化手法を組み合わせることで、より豊かに図形の特徴を捉えることができる