

データ解析で出会う統計的問題:

角度や時間など

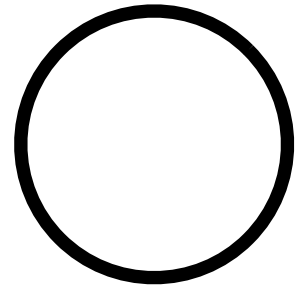
循環する変数の統計モデリング

生態学には  
circularなデータが  
たくさんある

粕谷英一      (九大・理・生物)  
かすや

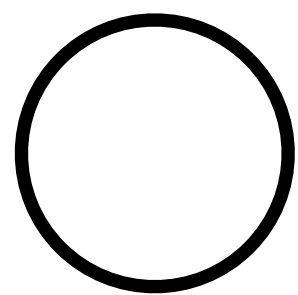


普通の量的なデータ 大小



circular

一周すると元に戻る



一周すると元に戻る

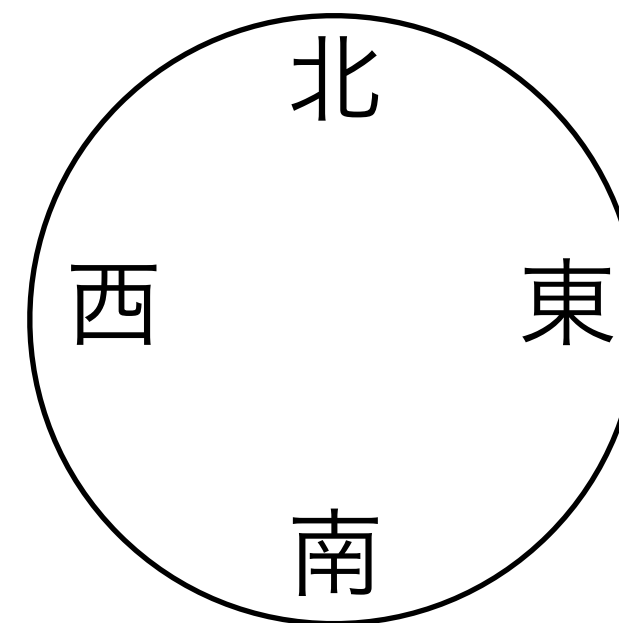
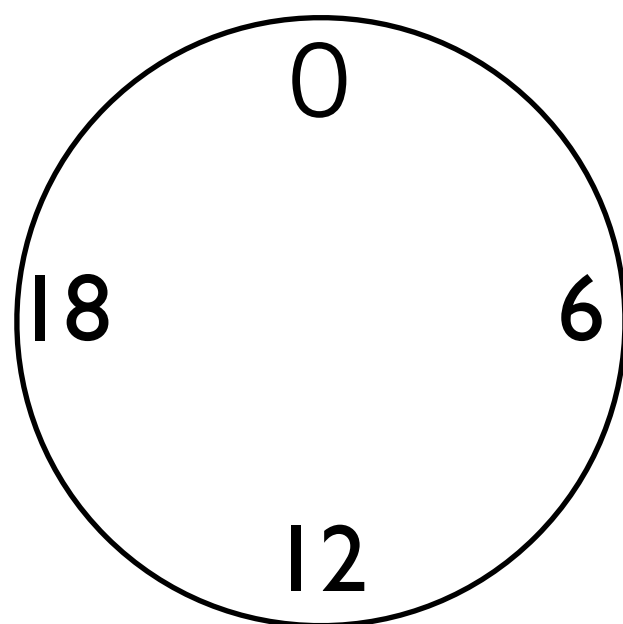
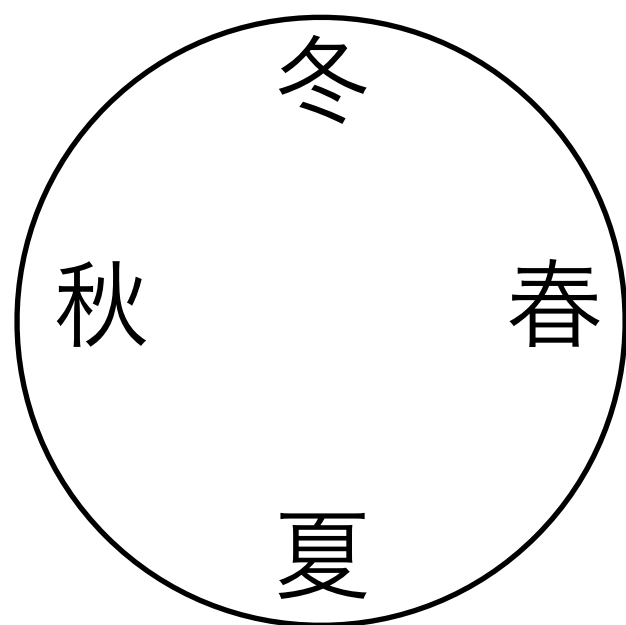
1 年の中の時期

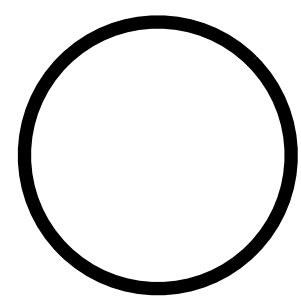
1 日の中の時刻

方位

方向

circular DNA





一周すると元に戻る

**360° 一周**

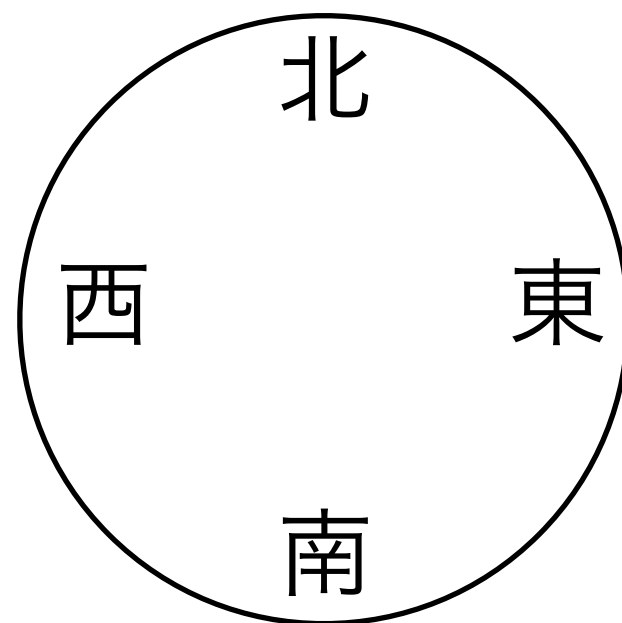
**$2\pi$  ラジアン 一周** 国際単位系(SI)

**24時間 一周**

**約365日 一周**

など

方位 方向



**circular** 一周すると戻ってくるから

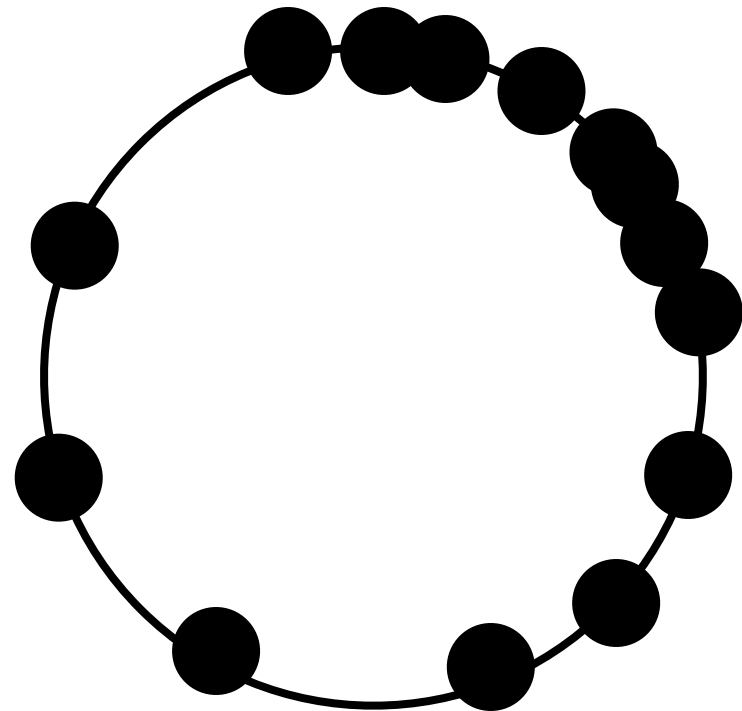
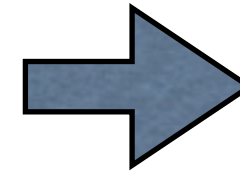
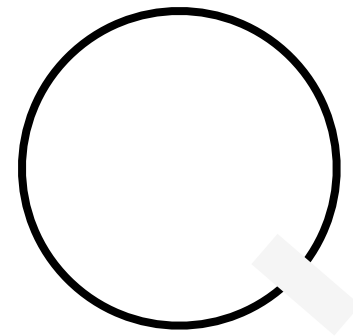
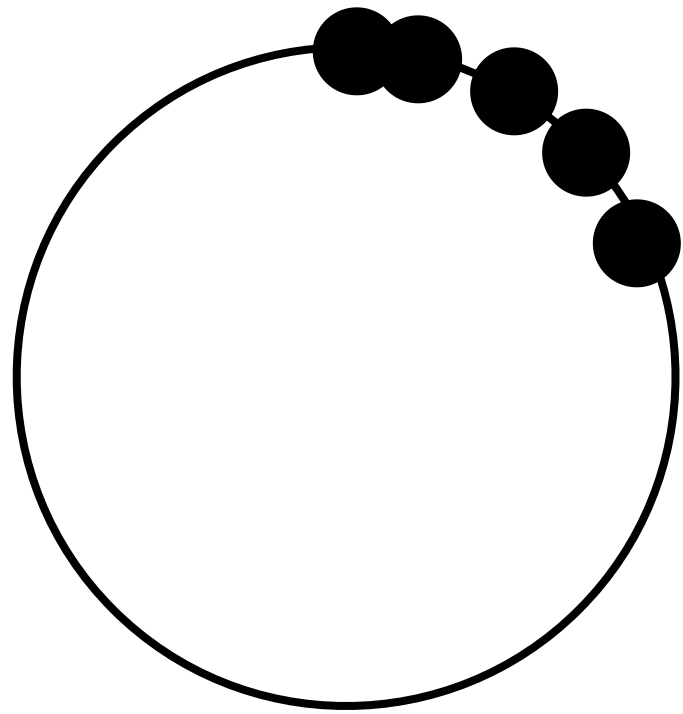
**angular** 角度だから

**directional** 方向だから

検索するときなどは実用的に重要

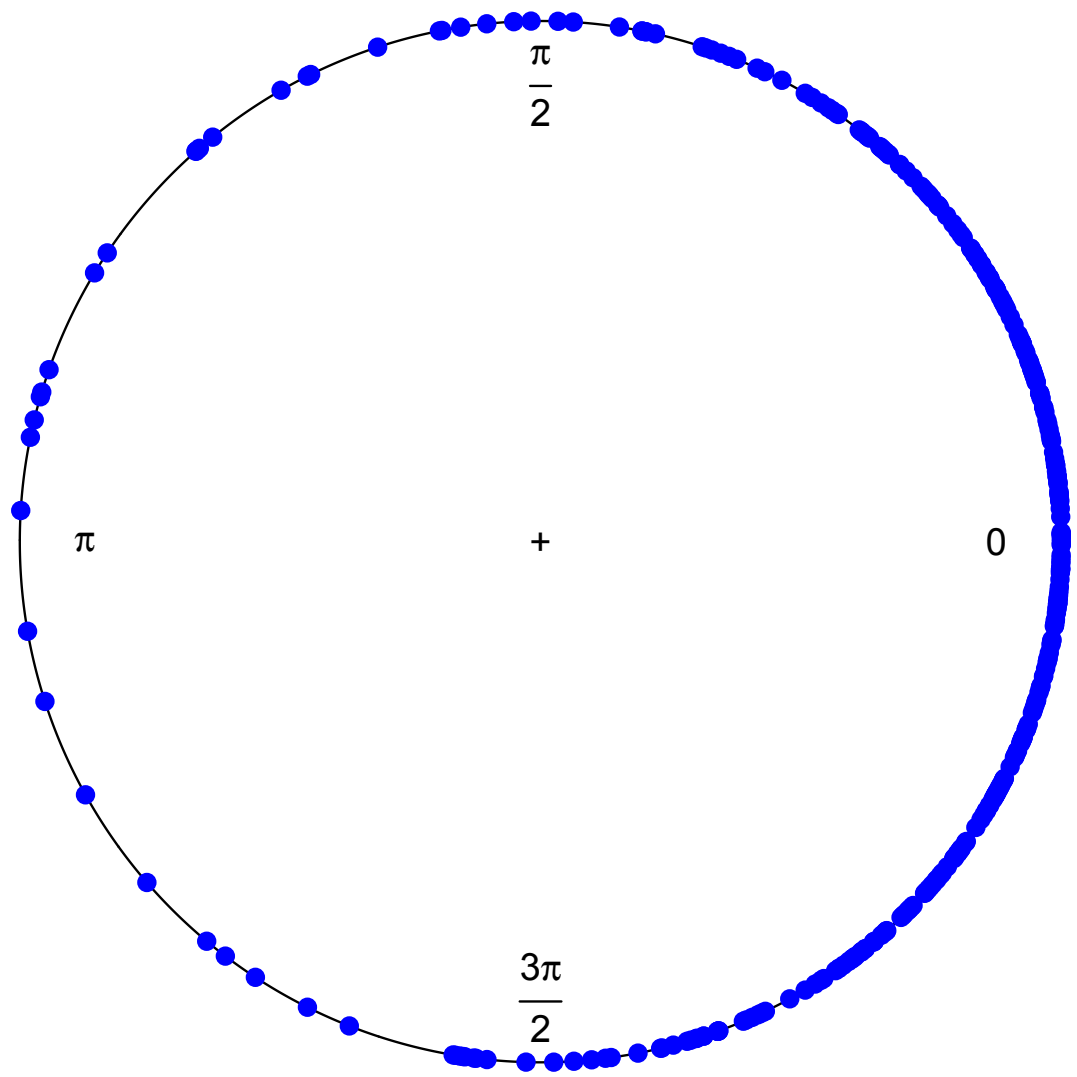


# 1 力所を切って開き、直線にする



データ点の数が少なく、ばらつきも小さいとき  
以外は行き詰まりやすい

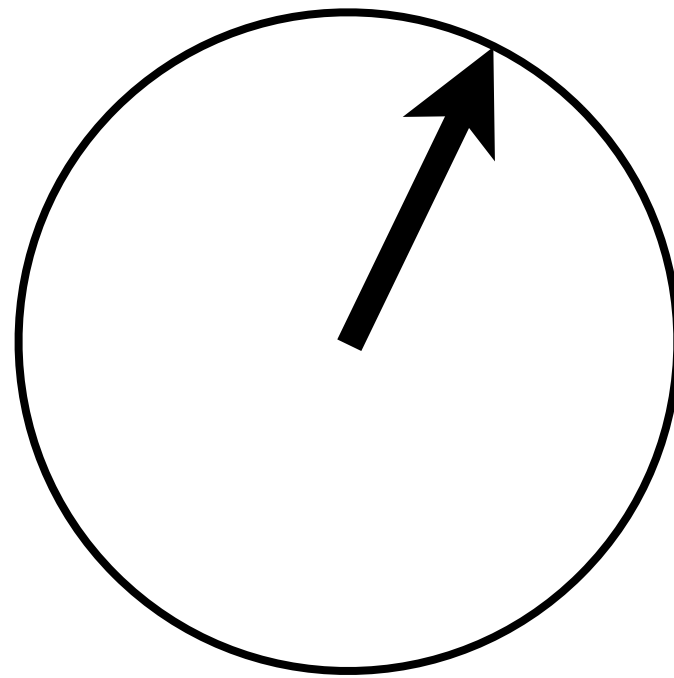
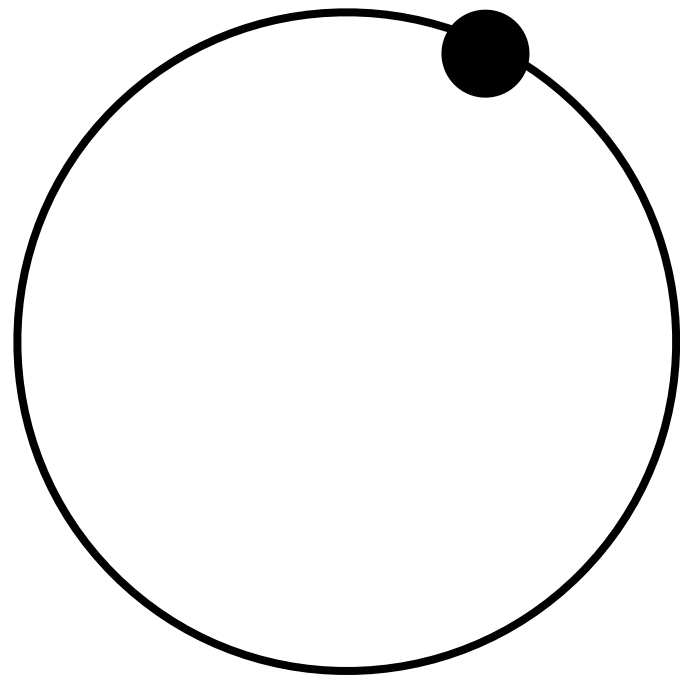
どこで”切り開くか”困る





# 扱い方

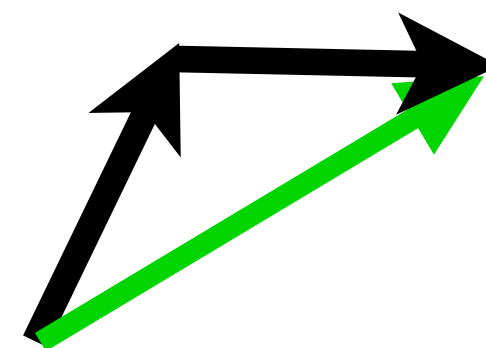
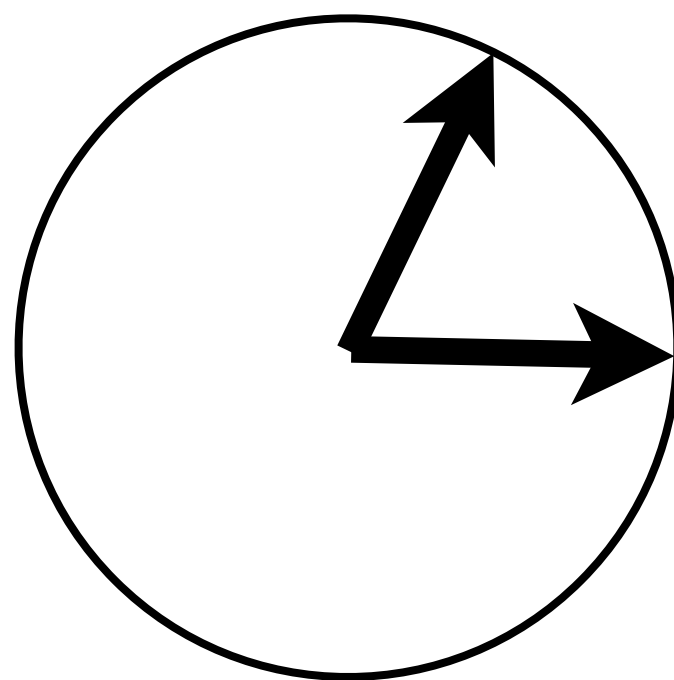
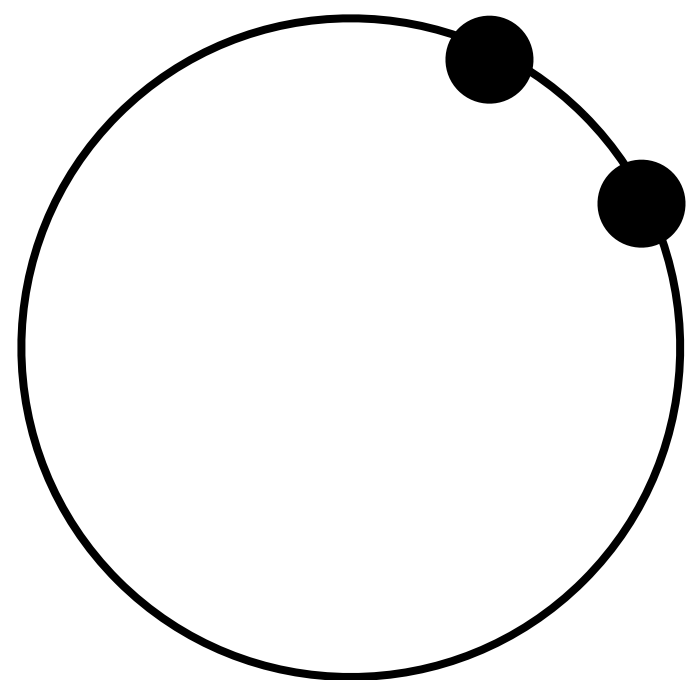
**平面上のベクトルだと考える**



# 扱い方 平面上のベクトルだと考える

合計

合成ベクトル

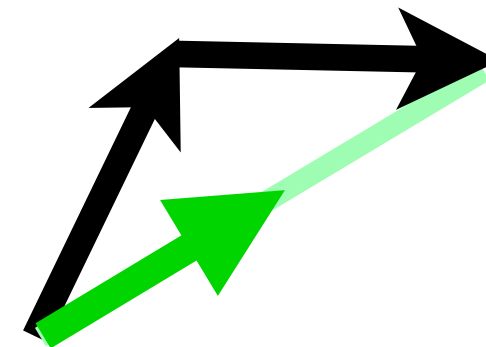
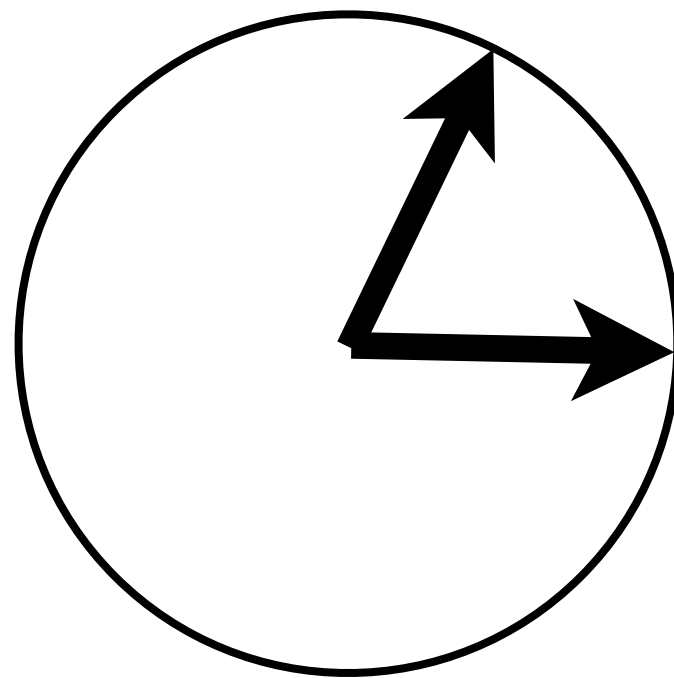
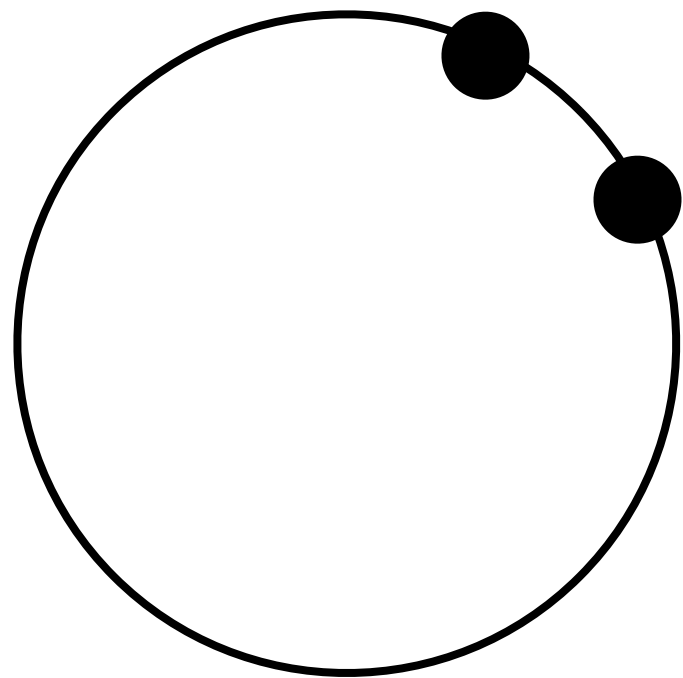


# 扱い方 平面上のベクトルだと考える

## 平均

## 平均合成ベクトルの方向

合成ベクトルの長さを(1/データの個数) 倍に

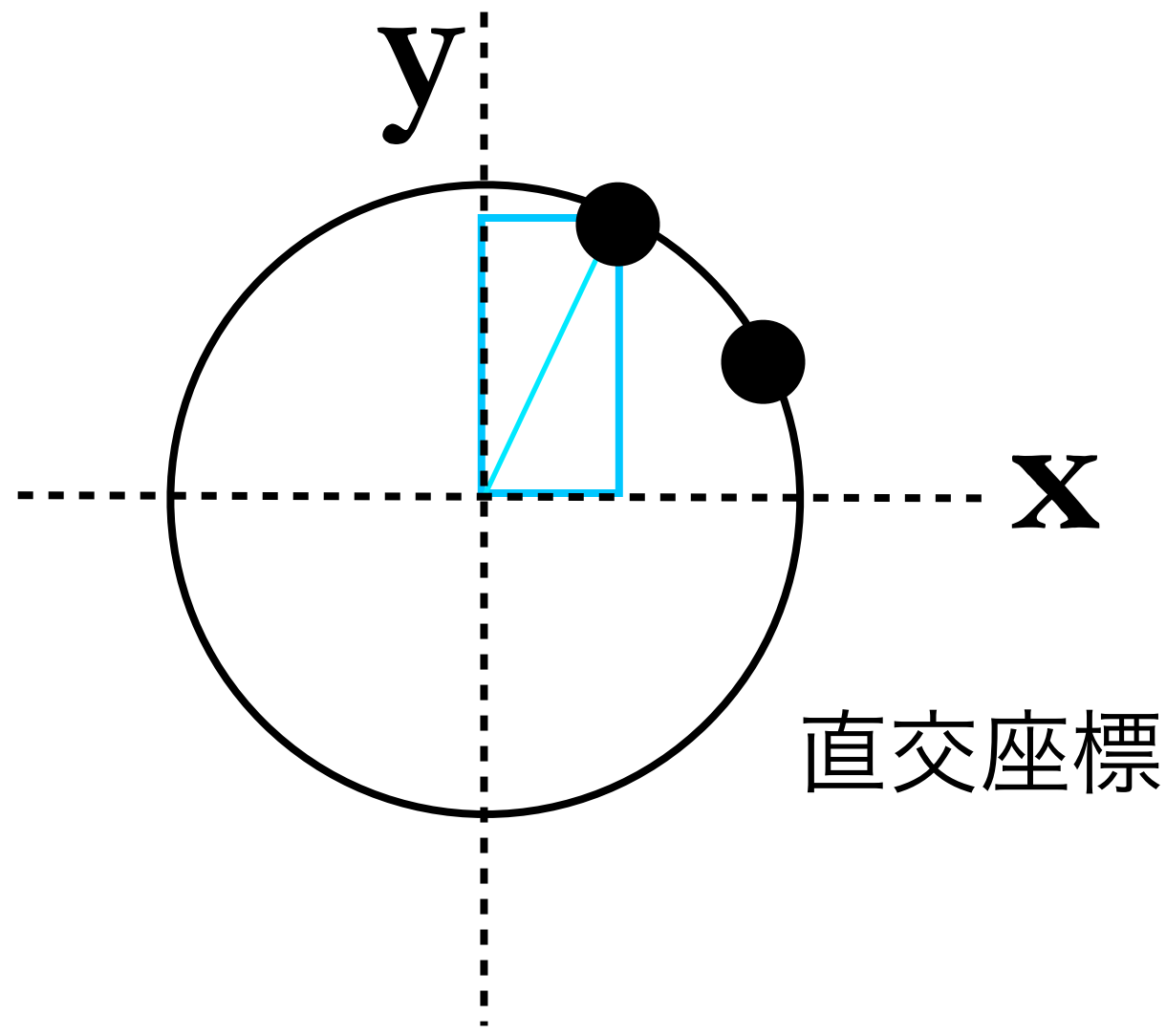


# 扱い方 平面上のベクトルだと考える 平均

合成ベクトル  
( $x$ の合計、 $y$ の合計)

長さを( $1/\text{データの個数}$ ) 倍に

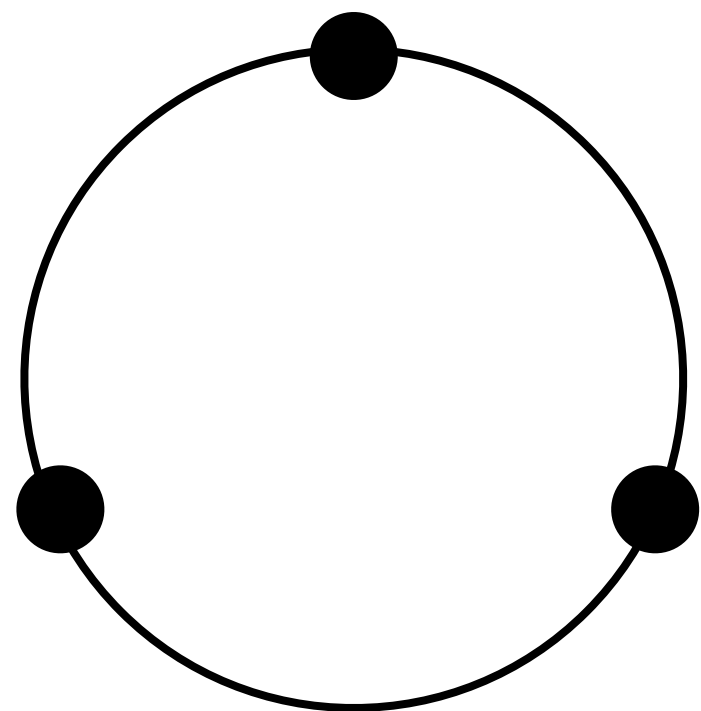
平均合成ベクトル



# 平均

# circularなデータ

特別な場合がある



合成ベクトルの長さがゼロ

平均がゼロであるというのとは異なる

# 平均

# circularなデータ

リニアな変数とちがう点

## 平均の平均

### 2つのデータセット

リニアな量

それぞれの平均

全体の平均

それぞれのデータの個数

circular

それぞれの平均

全体の平均

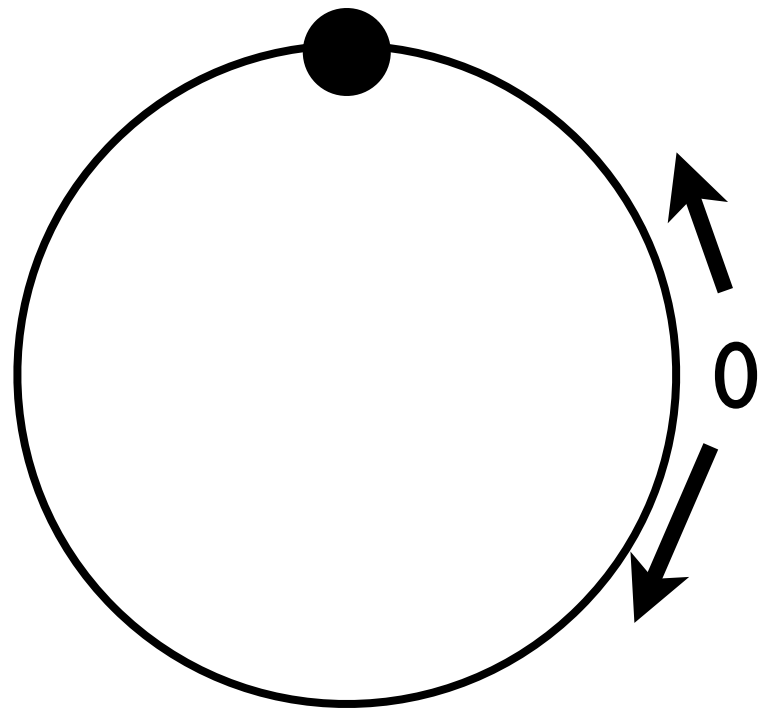
それぞれのデータの個数

▶ 合成ベクトル（平均合成ベクトルでもよい）の長さ



# circularなデータ

回転方向 (rotation)



数学などの角度：反時計回り

平均合成ベクトル

ばらつきが大きいと短い

**ばらつきの指標**

circular variance=

1－（平均合成ベクトルの長さ）

全データがみな同じ（ベクトルが同じ方向）のときゼロになる

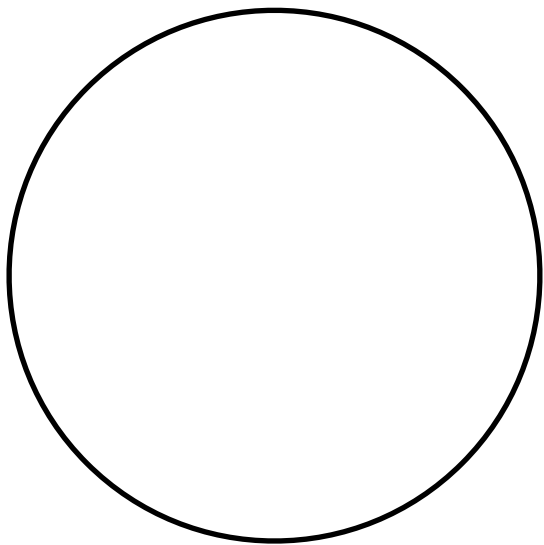


## circular median

円周を半分

データの個数が等しい

データの多数が、ちょうど反対側よりもその点に近い



# ばらつきの指標

circular dispersion

$$\delta = \frac{1 - R_2}{2R^2}$$

$R$ =平均合成ベクトル長

$R_2$ =角度を2倍にしたベクトルの平均合成ベクトル長

circular standard deviation

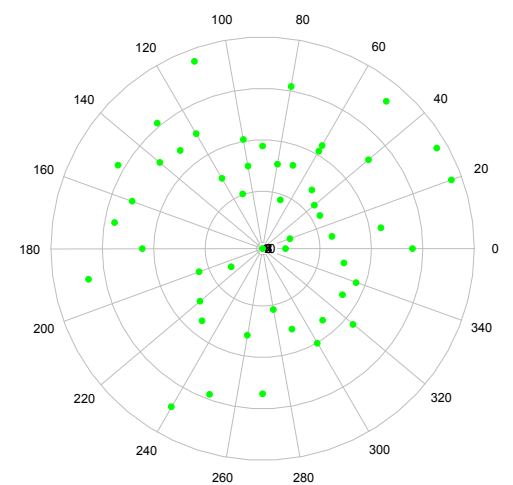
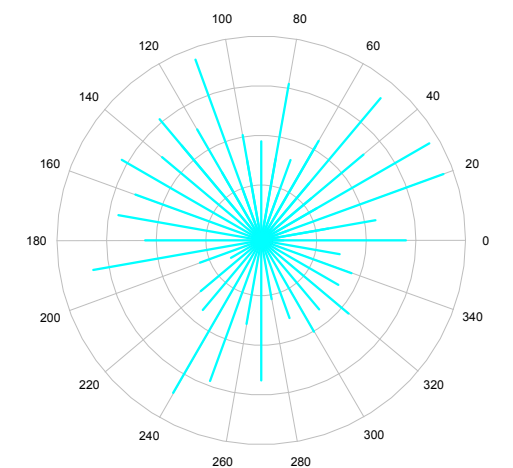
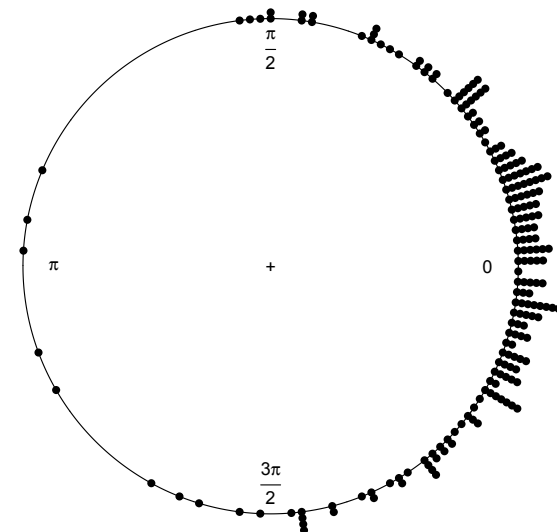
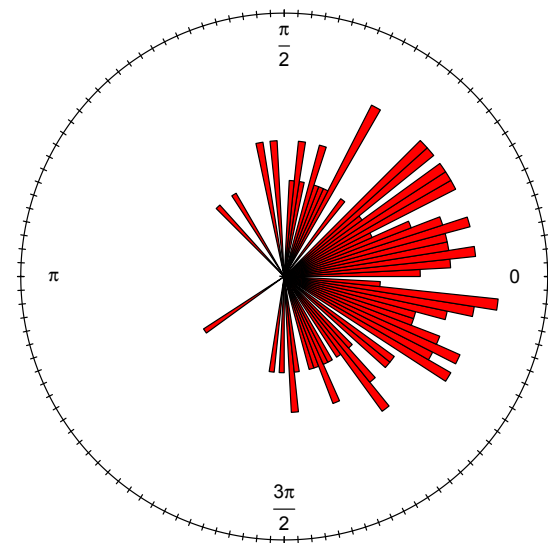
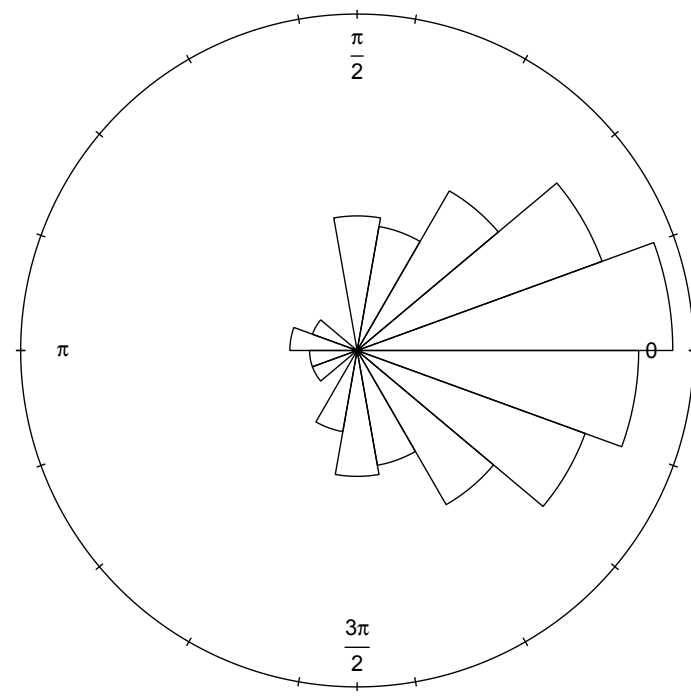
$\{-2\log(\text{平均合成ベクトルの長さ})\}$ の平方根

circular range

すべてのデータを含む円周上の弧のうち最短のものの長さ

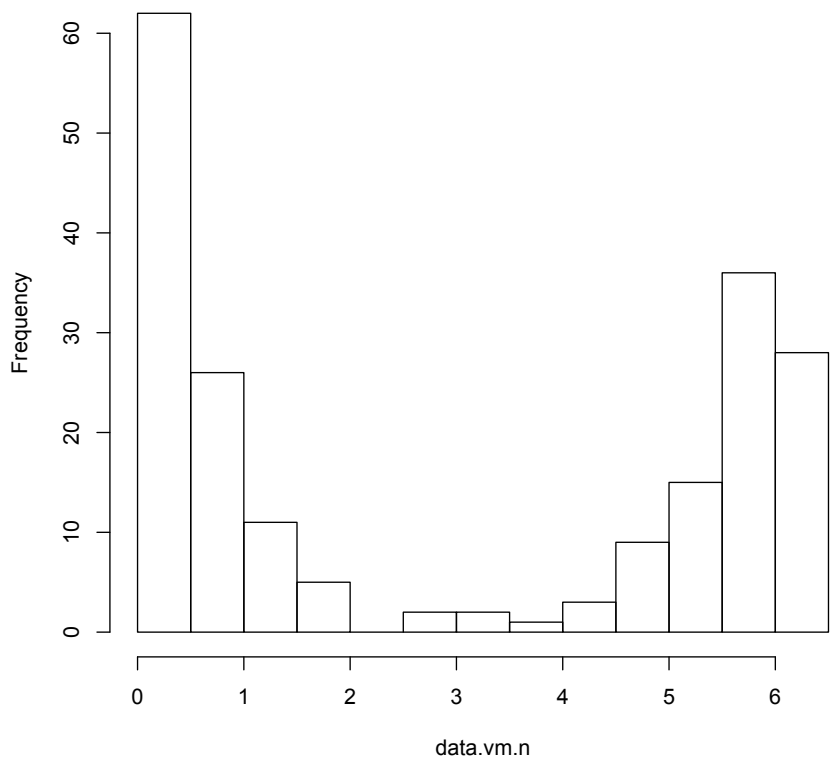
# 頻度分布のグラフ

ヒストグラムなどにあたる

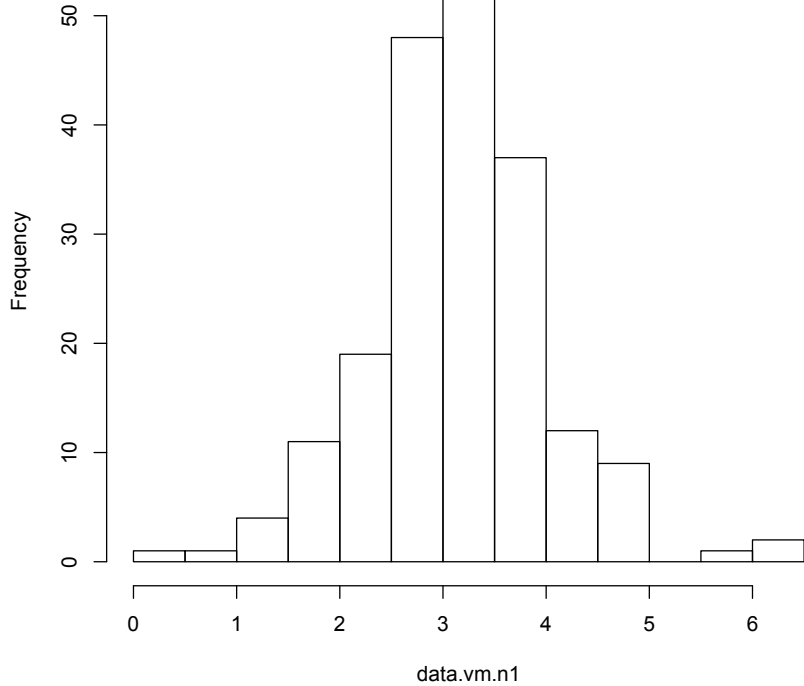


rose diagram

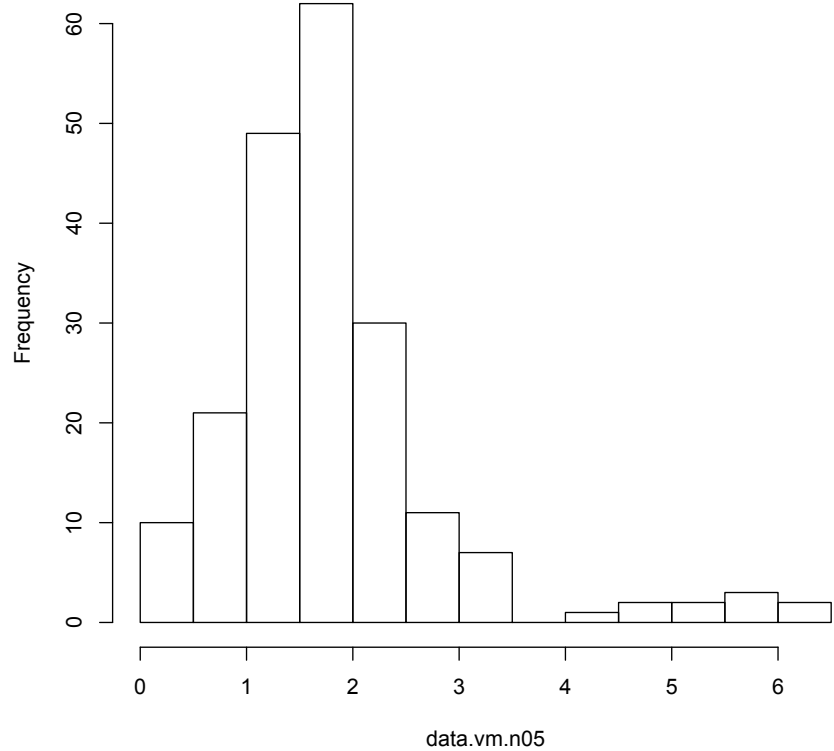
Histogram of data.vm.n



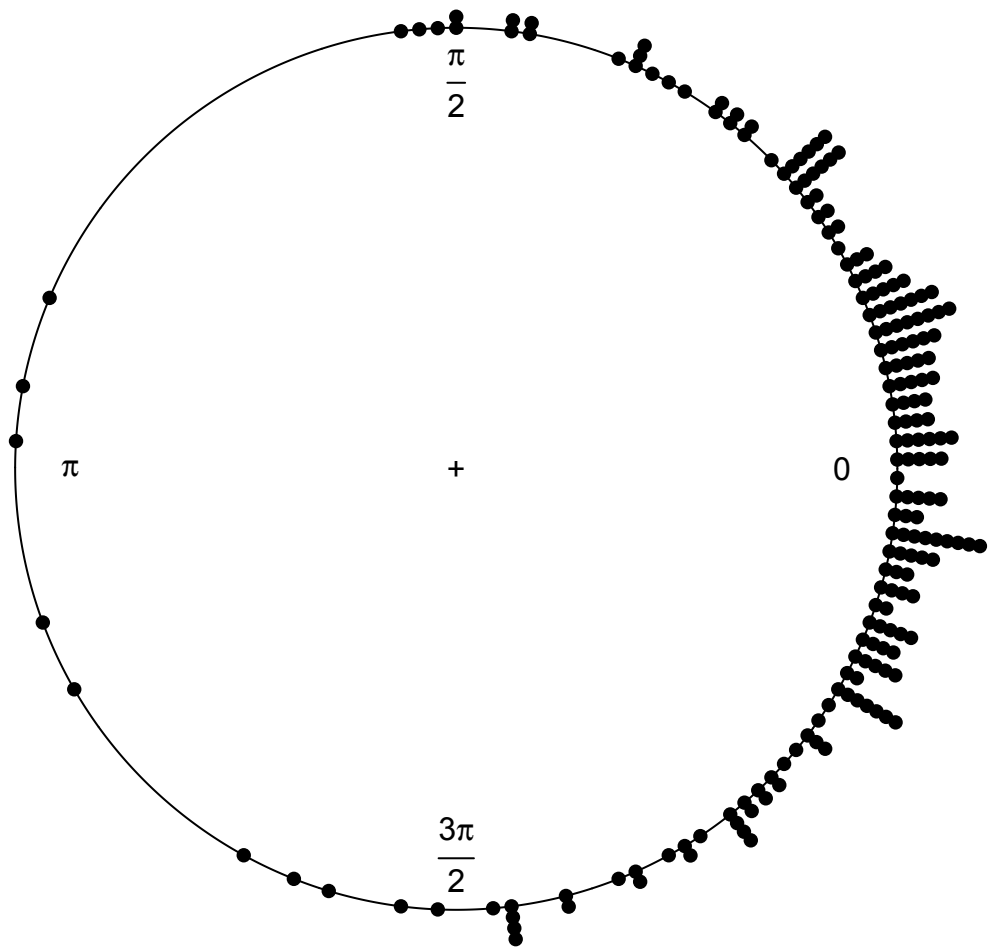
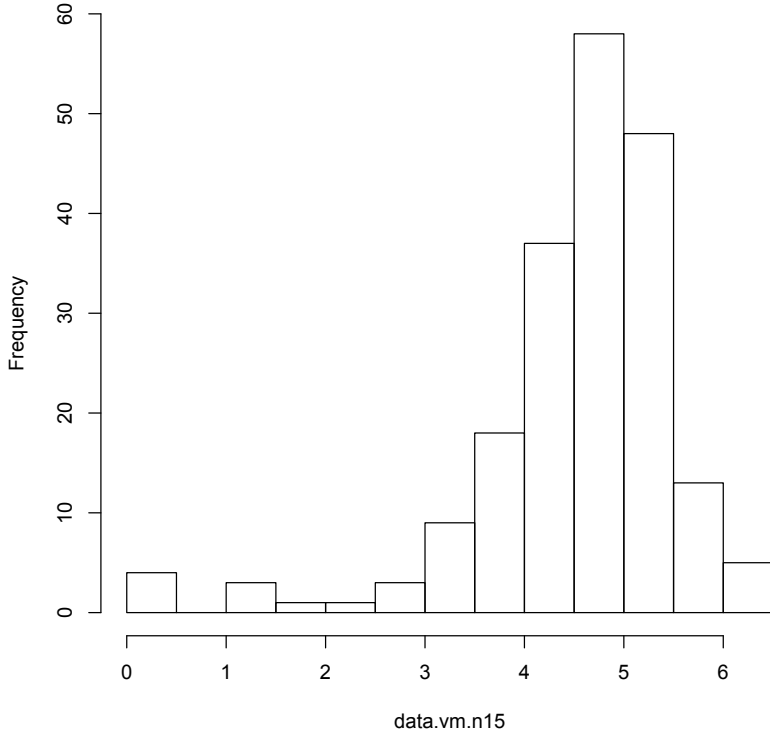
Histogram of data.vm.n1



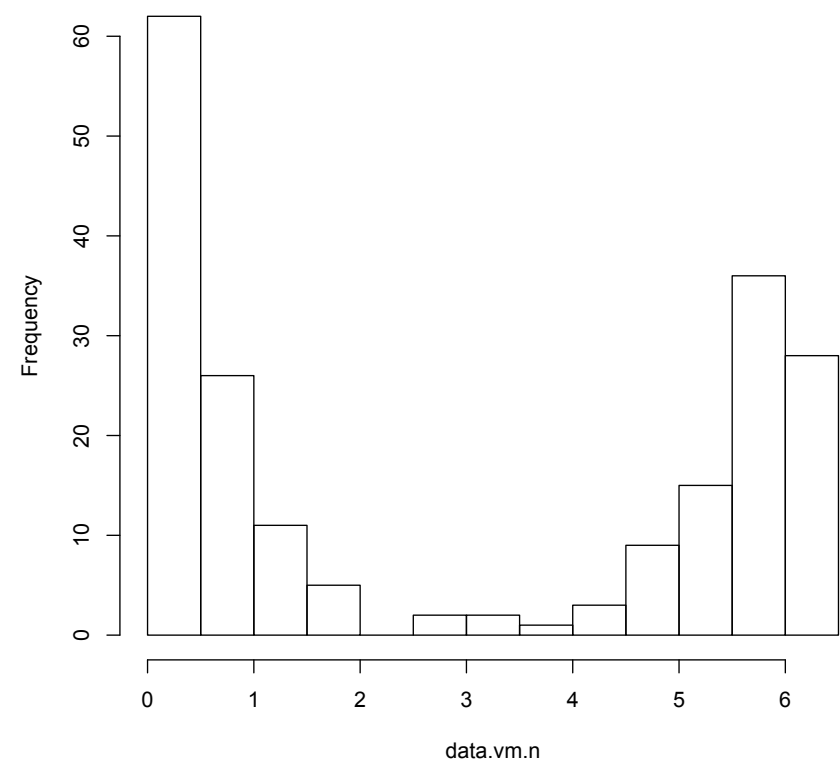
Histogram of data.vm.n05



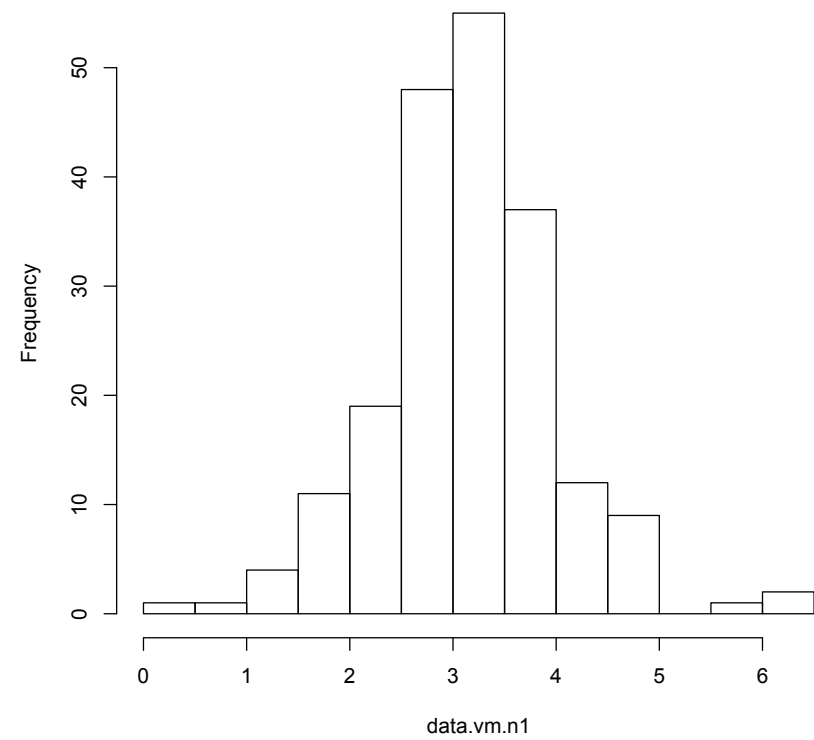
Histogram of data.vm.n15



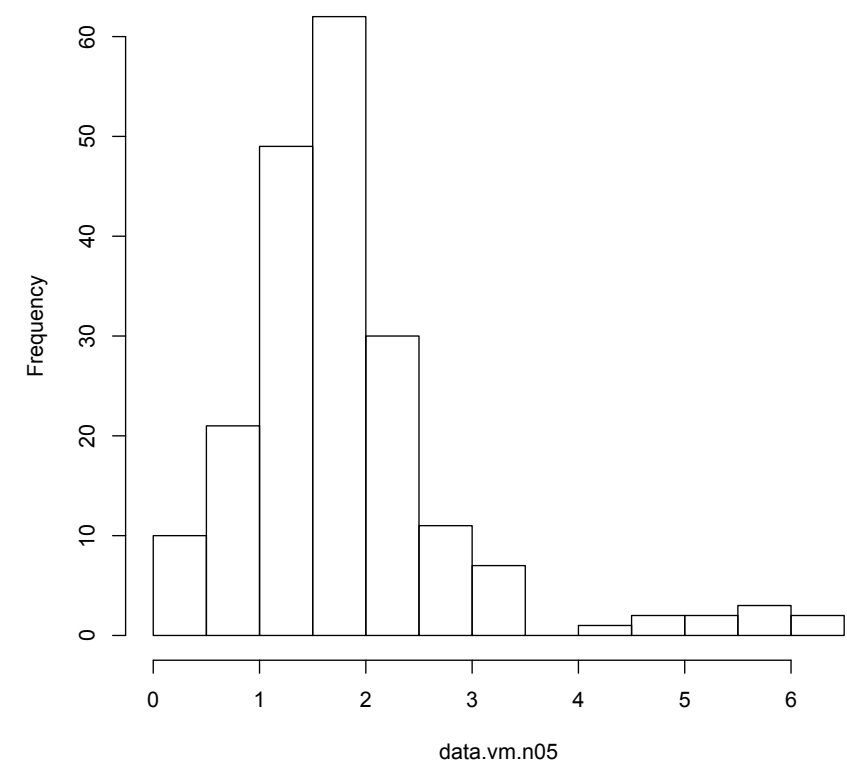
Histogram of data.vm.n



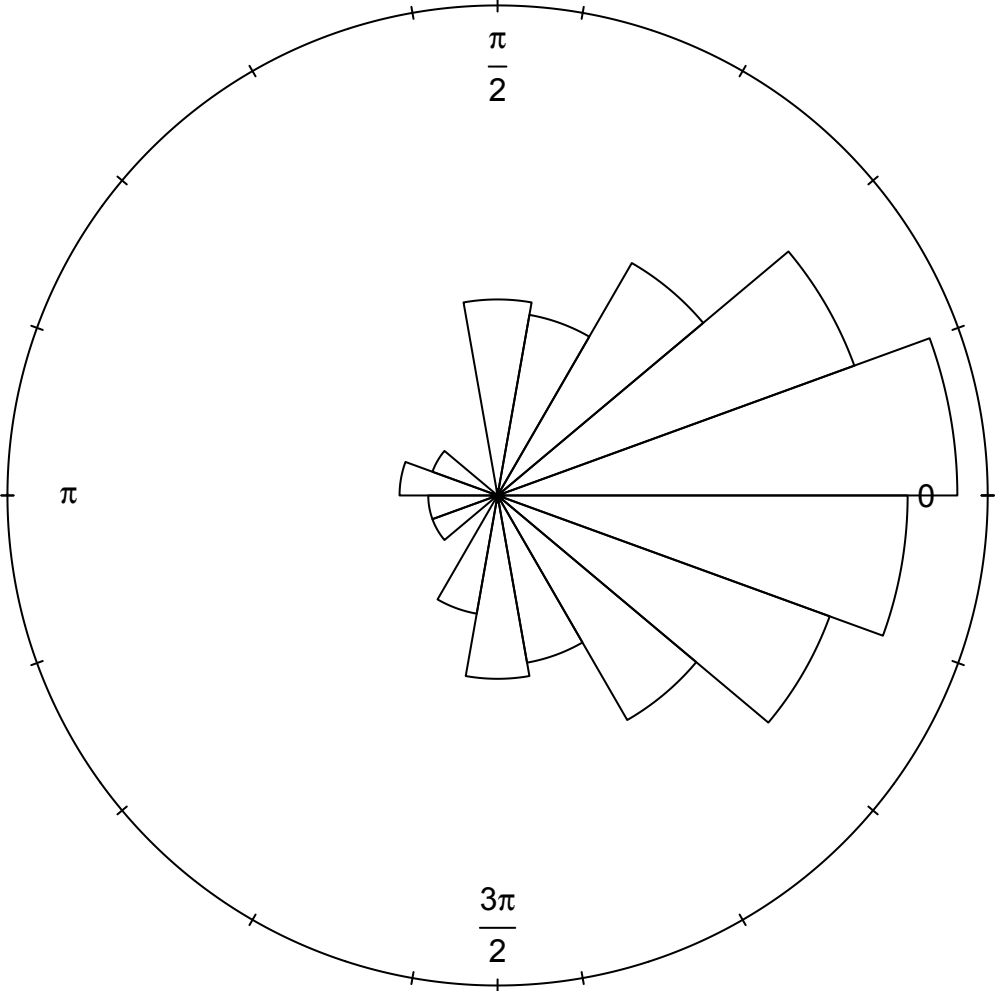
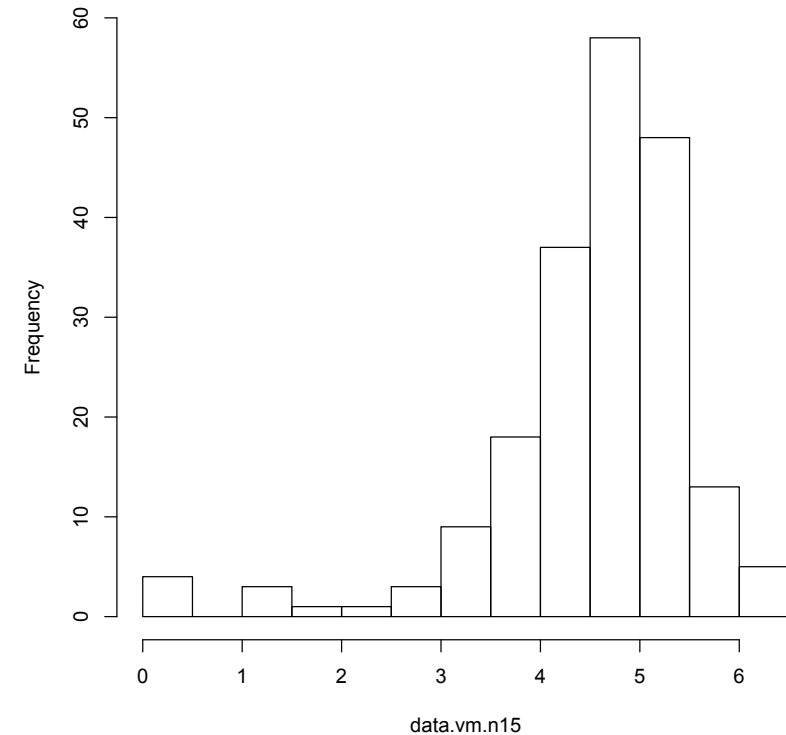
Histogram of data.vm.n1



Histogram of data.vm.n05

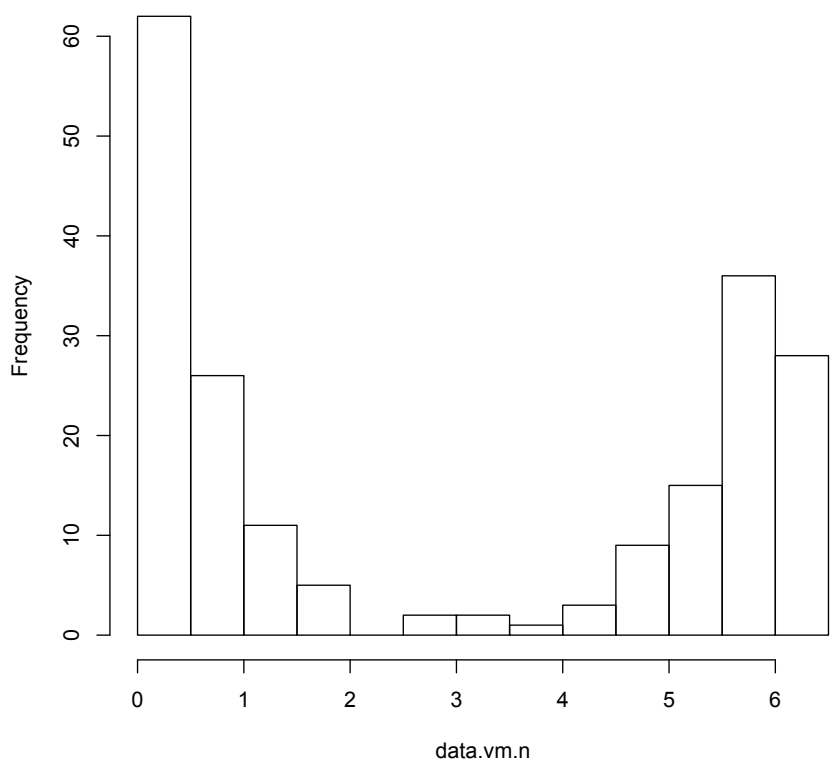


Histogram of data.vm.n15

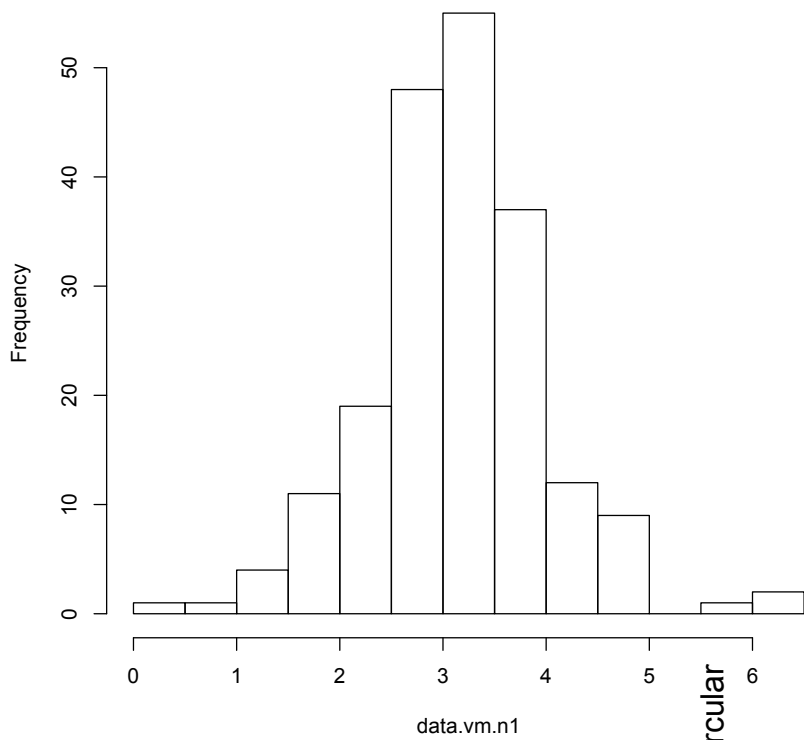


rose diagram

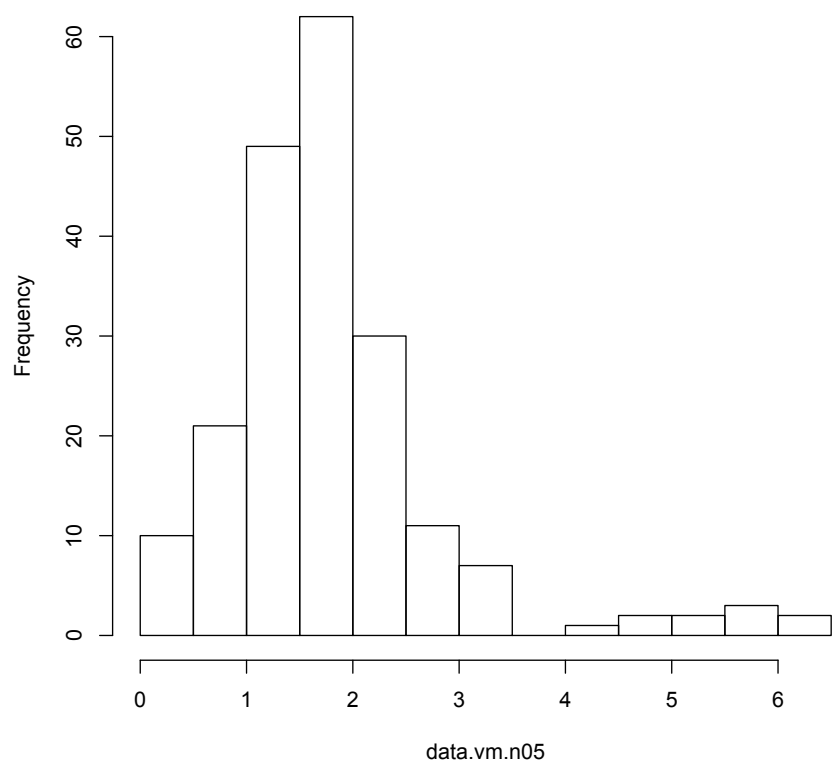
Histogram of data.vm.n



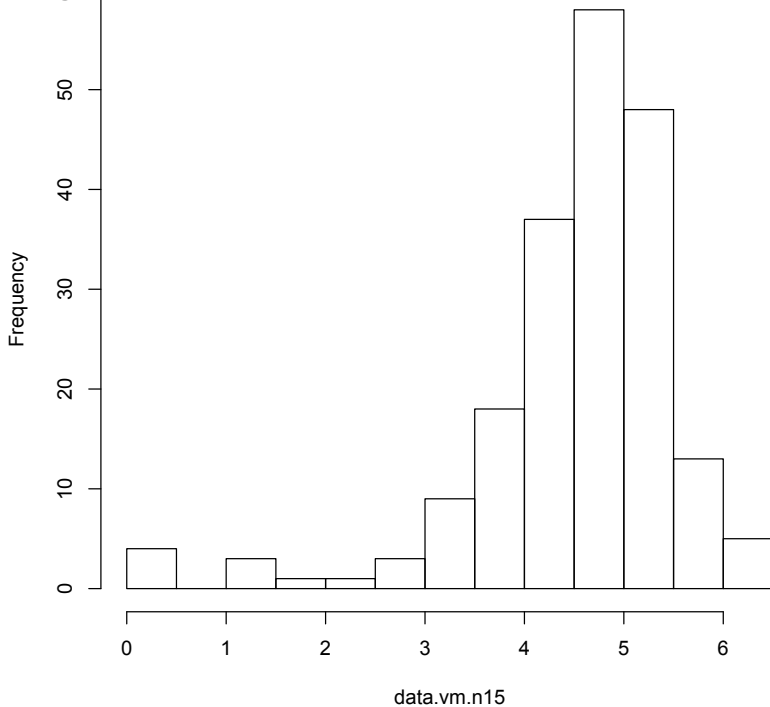
Histogram of data.vm.n1



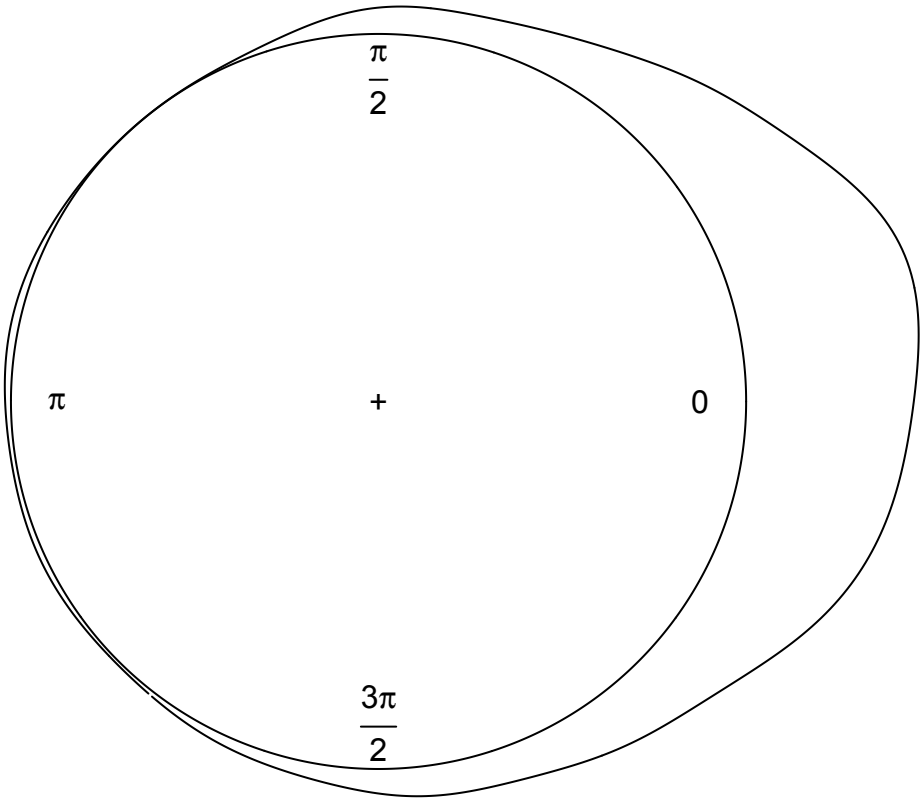
Histogram of data.vm.n05



Histogram of data.vm.n15



Density circular



kernel densityプロット

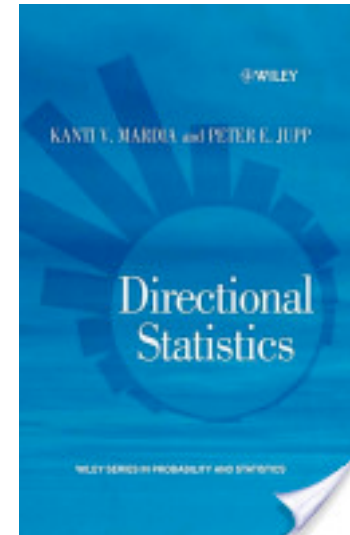


# circularなデータの解析を主なテーマとした本

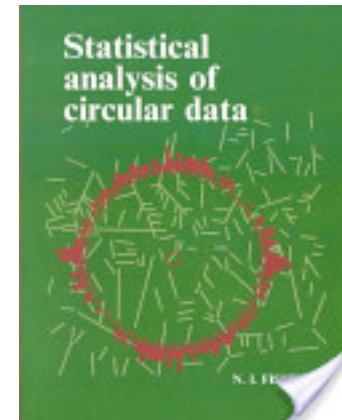
(”6冊だけ” Pewseyら2014)



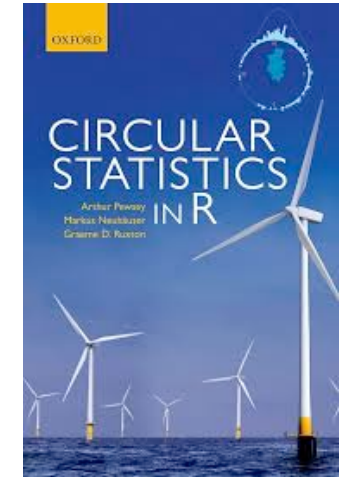
Mardia(1972)  
*Statistics in Directional Data*



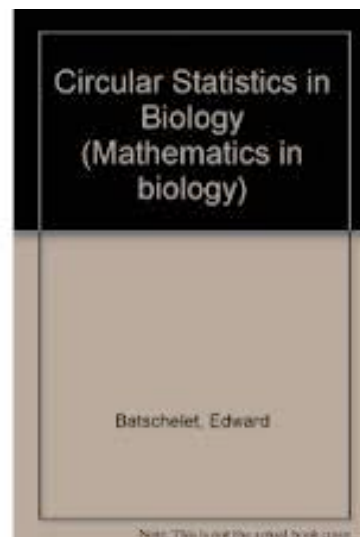
Mardia&Jupp(2000)  
*Directional Statistics*



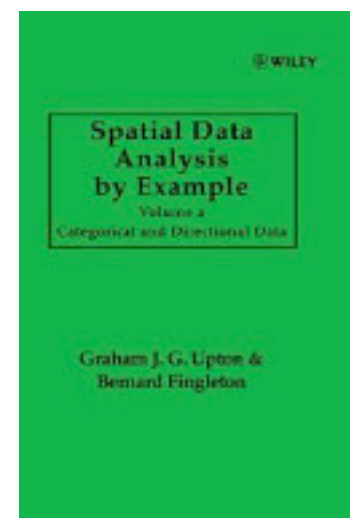
Fisher(1993)  
*Statistical Analysis of Circular Data*



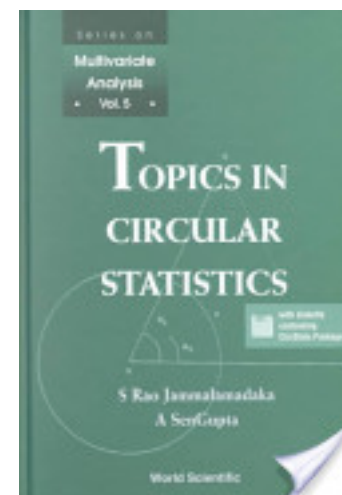
Circular Statistics in R  
Pewseyほか(2014)



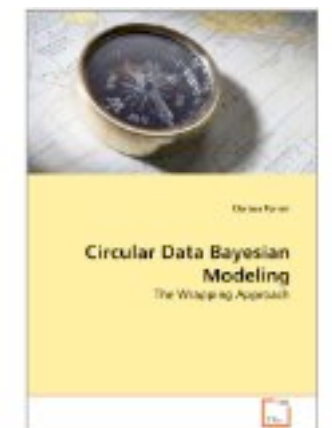
Batschelet(1981)  
*Circular statistics in biology*



Upton&Fingleton(1989)  
*Spatial Data Analysis by Example2*



Jammalamadaka&Sengupta(2001)  
*Topics in Circular Statistics*



Ferrari(2010)  
*Circular Data Bayesian Modeling*

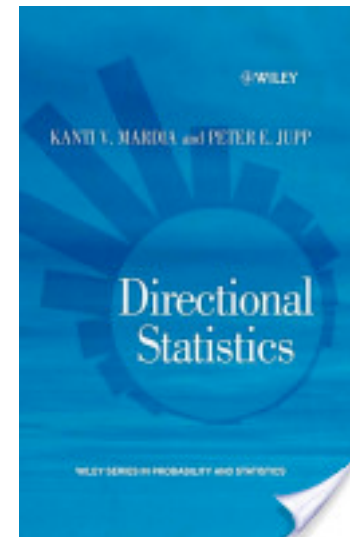
# circularなデータの解析を主なテーマとした本

Zar

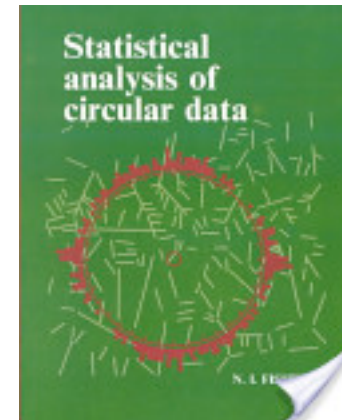
*Biostatistical Analysis*



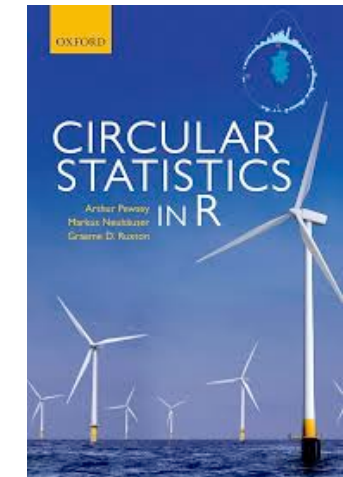
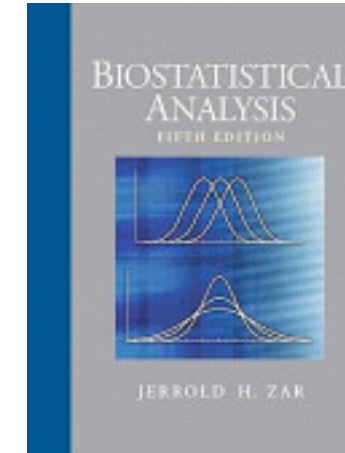
Mardia(1972)  
*Statistics in Directional Data*



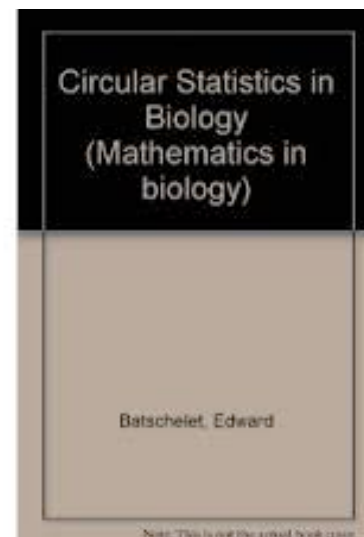
Mardia&Jupp(2000)  
*Directional Statistics*



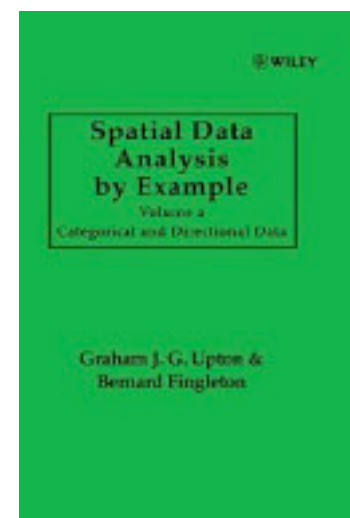
Fisher(1993)  
*Statistical Analysis of Circular Data*



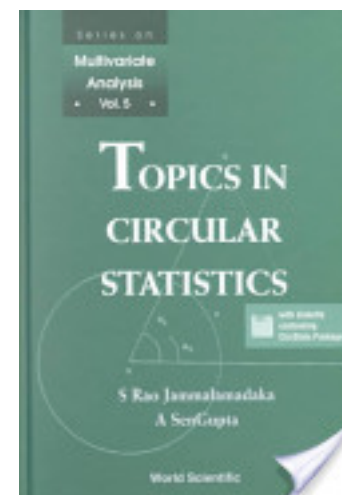
Circular Statistics in R  
Pewseyほか(2014)



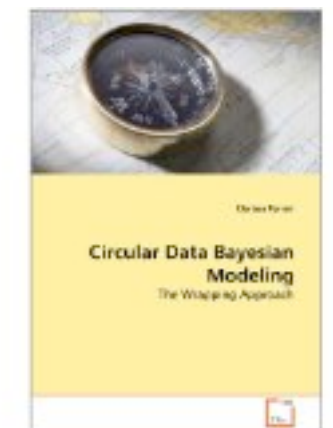
Batschelet(1981)  
*Circular statistics in biology*



Upton&Fingleton(1989)  
Spatial Data Analysis bei Example2



Jammalamadaka&Sengupta(2001)  
Topics in *Circular Statistics*



Ferrari(2010)  
Circular Data Bayesian Modeling



# 相関

**circular-linear**

**circular-circular**

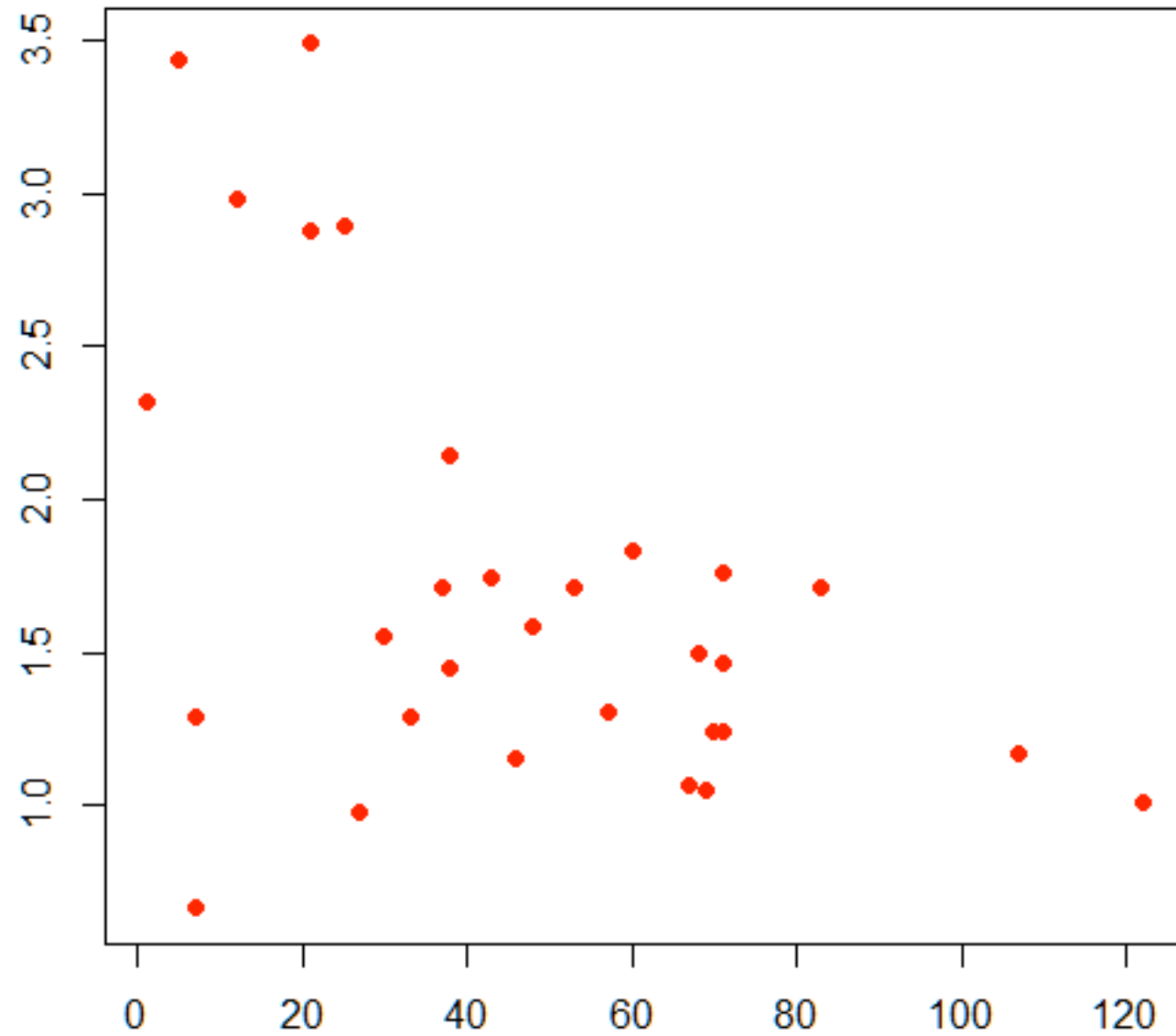
linear-linear

それぞれ専用の相関係数など

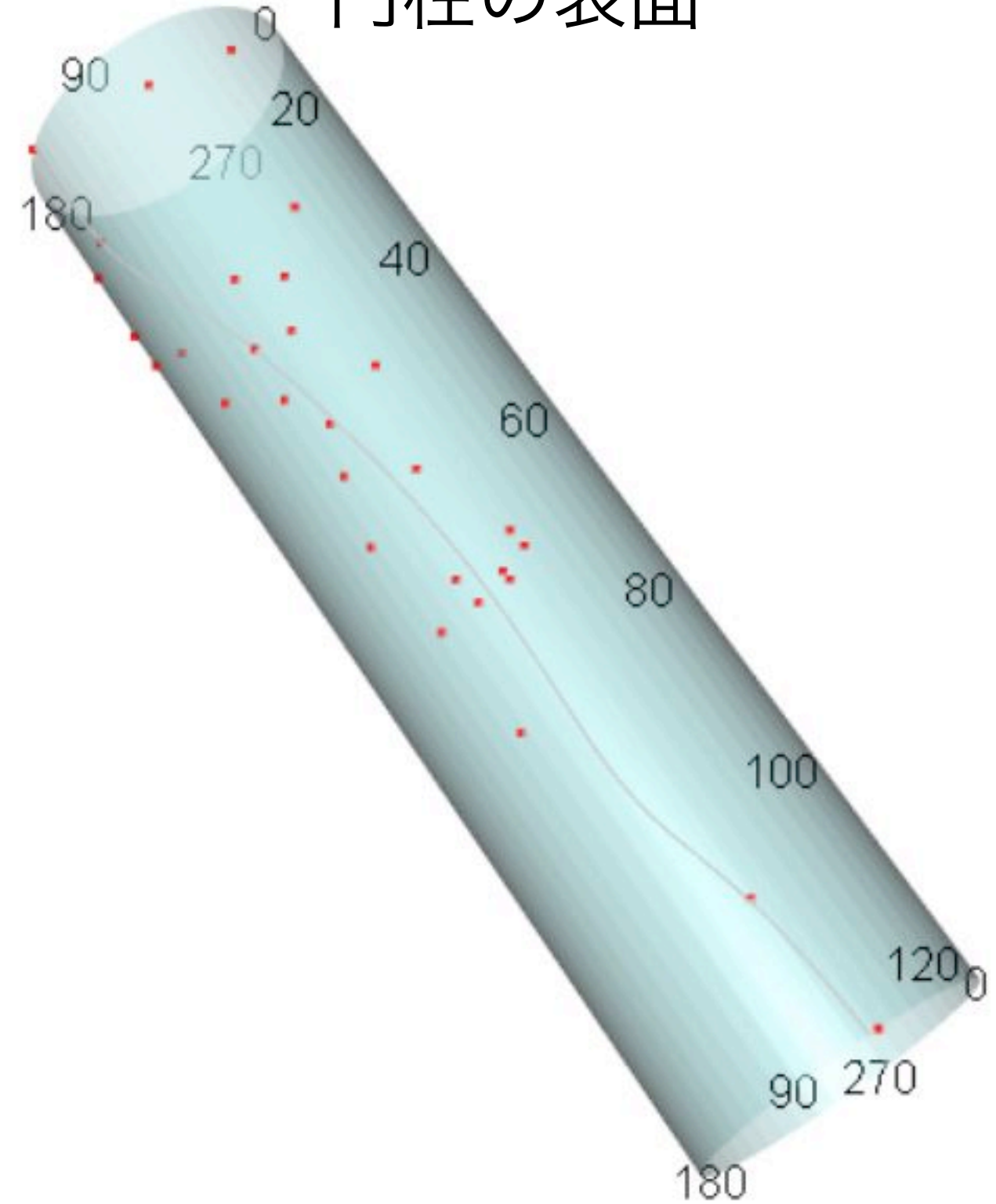
# 散布図

# circular-linear

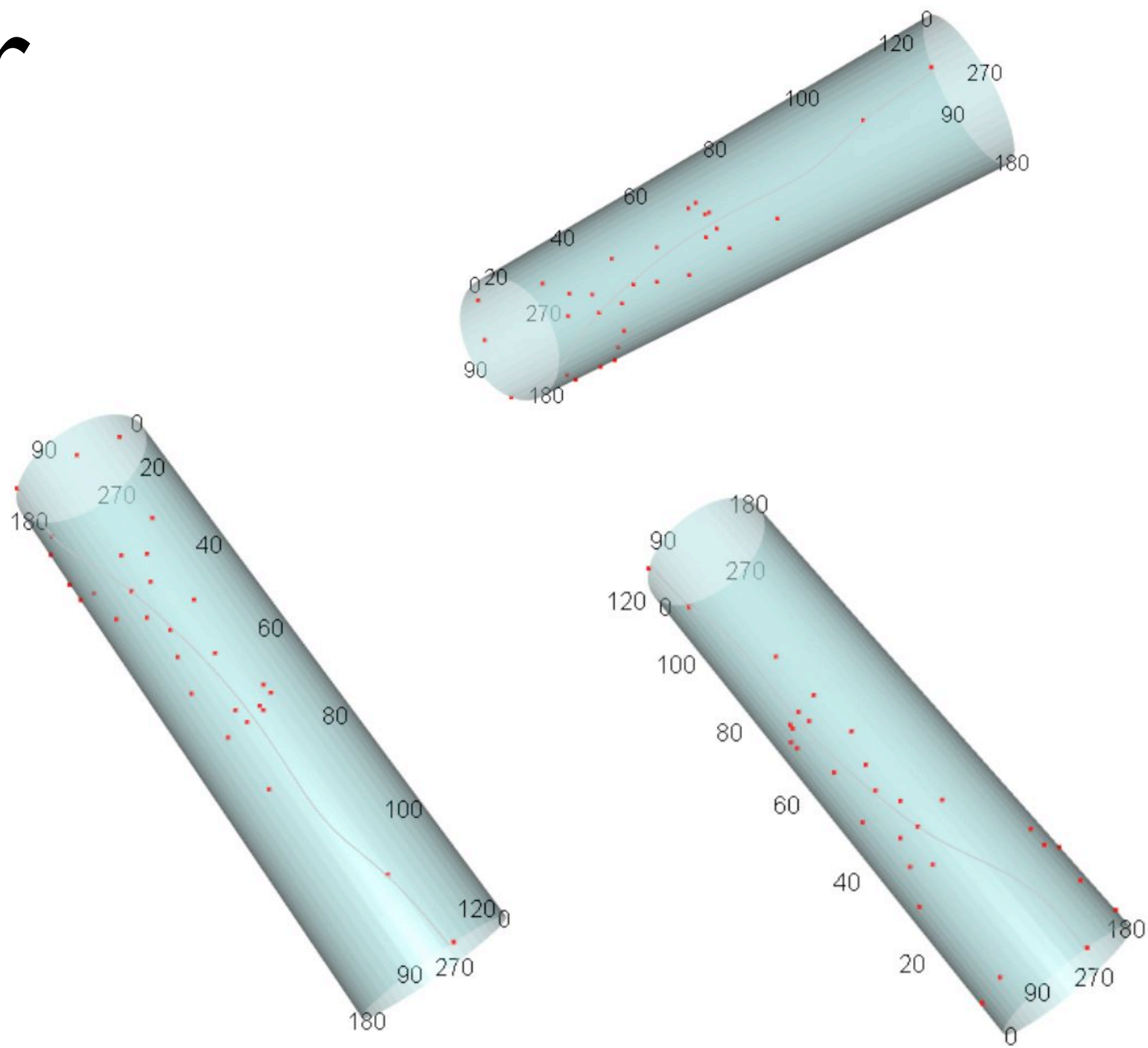
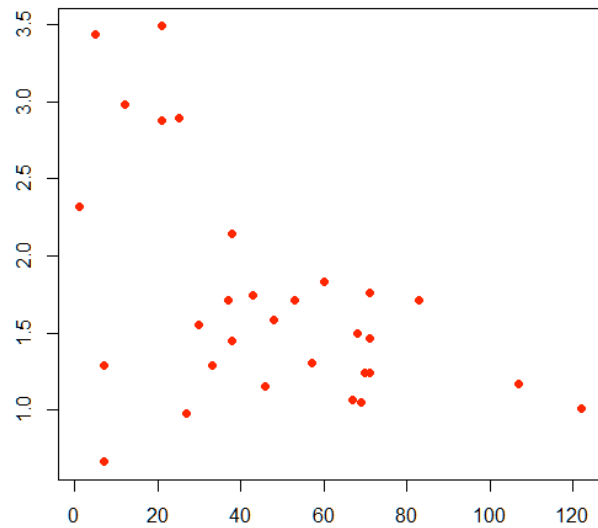
円 直線



円柱の表面



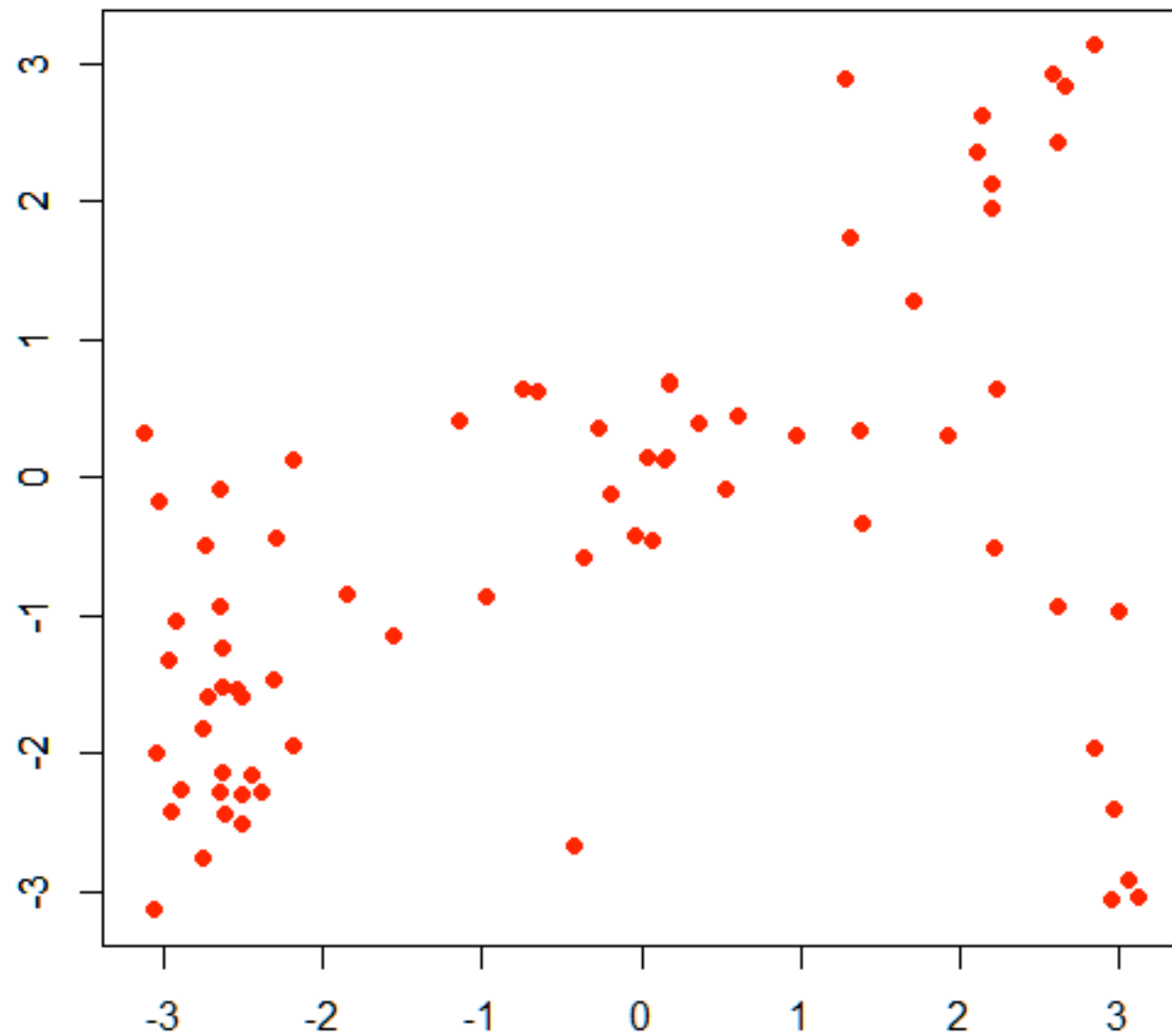
# circular-linear



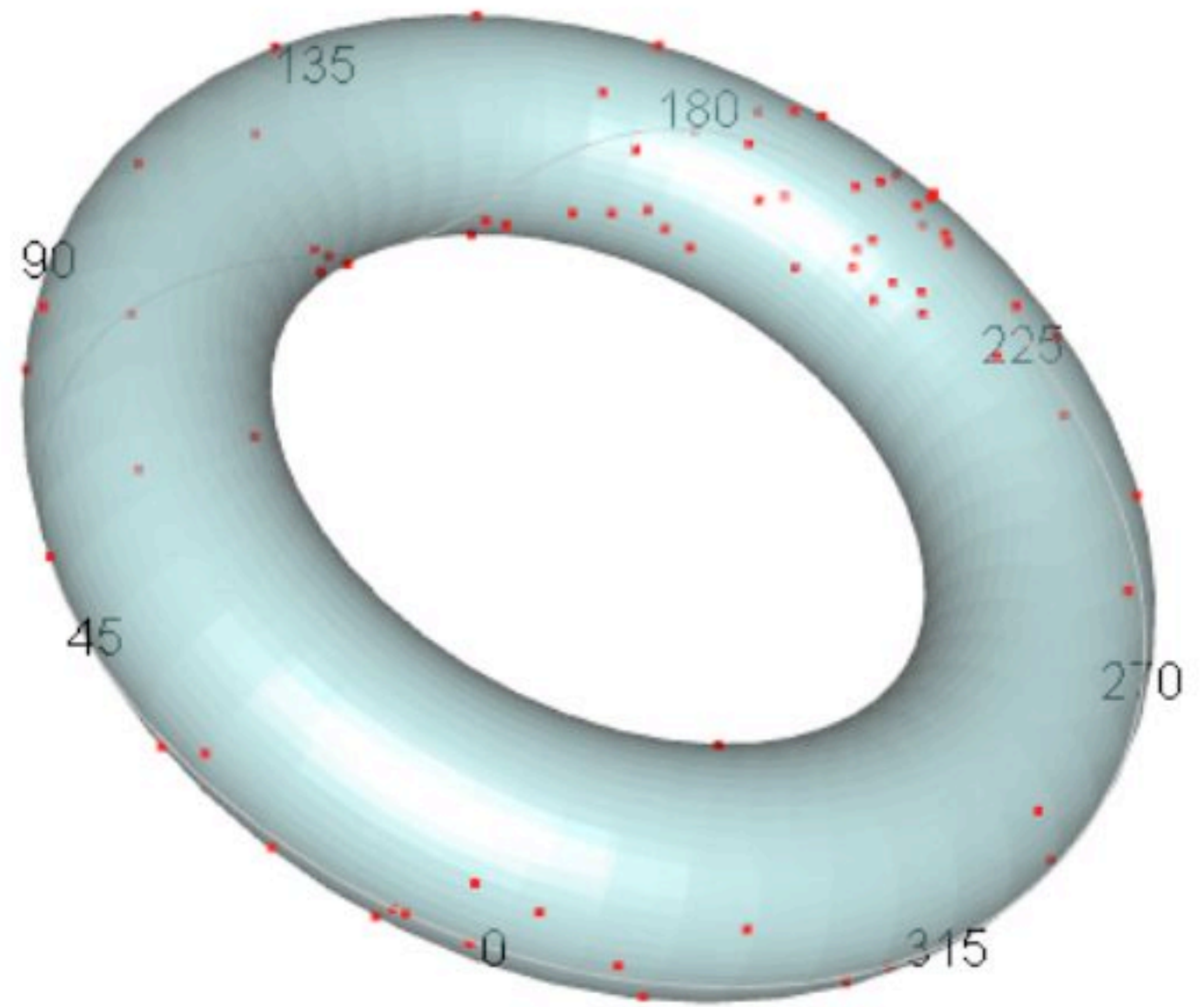
# circular-circular

円

円



トーラス（ドーナツ状）の表面

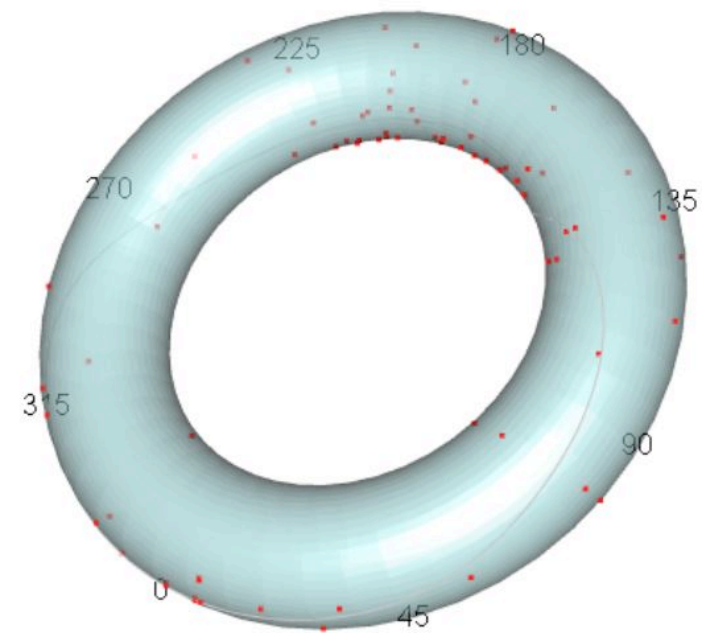
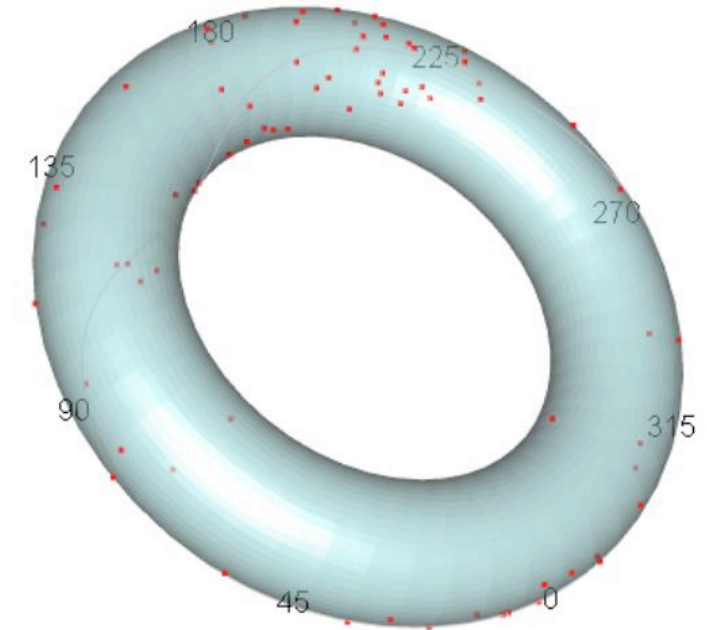
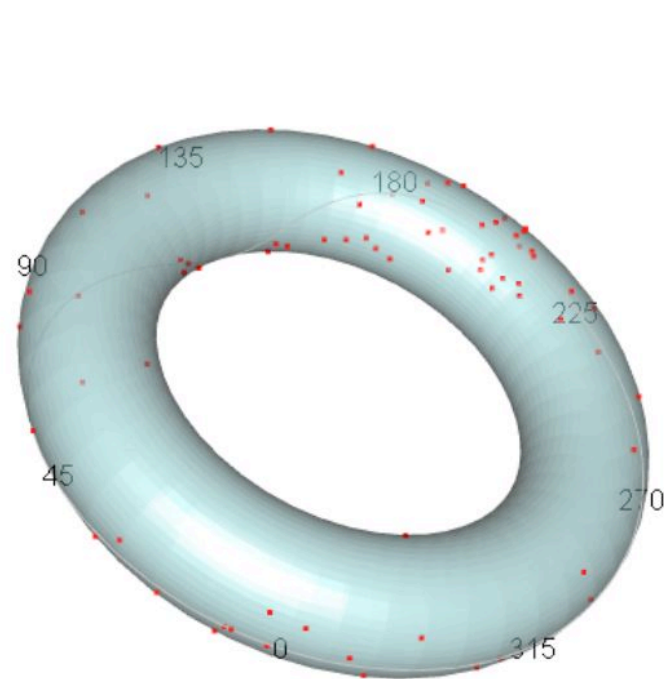
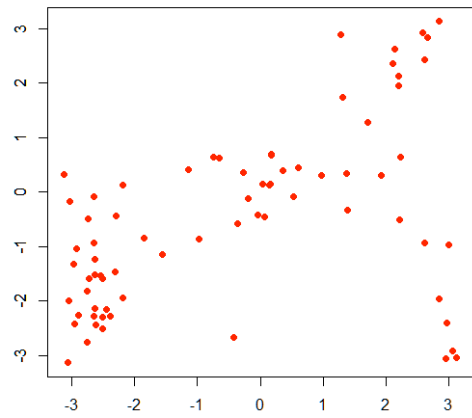


# circular-circular

円

円

トーラス（ドーナツ状）の表面





# 分布

一様分布

von Mises分布

ラッピング

# 一様分布

帰無仮説的な役割

例. Rayleighの検定

一様分布が帰無仮説

平均合成ベクトル長が検定統計量

平均合成ベクトル長がゼロになるようなパターンは検出できない  
一山でない分布の検出には向かない



# von Mises分布

確率密度関数

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

平均  $\mu$

分散の指標  $\kappa$  (ただし大きい方がばらつきが小さい)

# von Mises分布

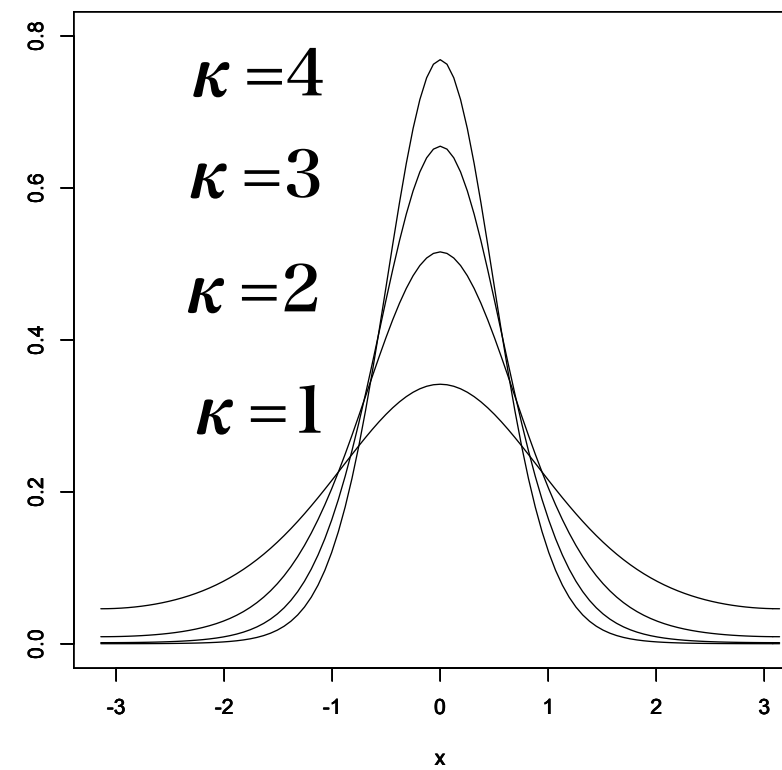
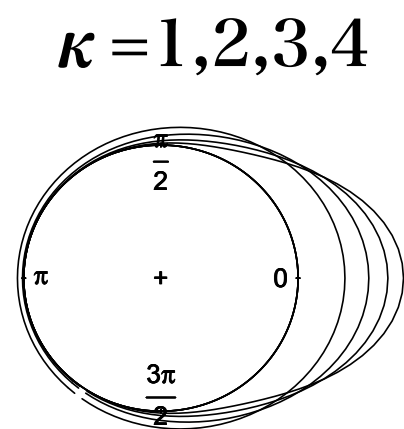
$$f(x) = \frac{e^{\kappa \cos(x-\mu)}}{2\pi I_0(\kappa)} \quad \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

正規分布と比べてみる

平均  $\mu$

分散の指標  $\kappa$  (ただし大きい方がばらつきが小さい)

# von Mises分布

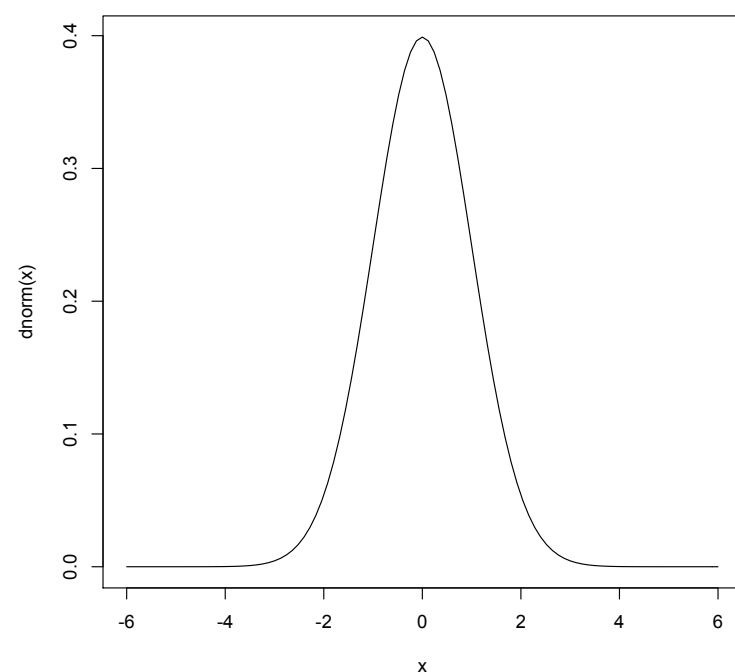


# ラッピング

リニアな変数の確率分布

長方形のラップ（あるいは紙）

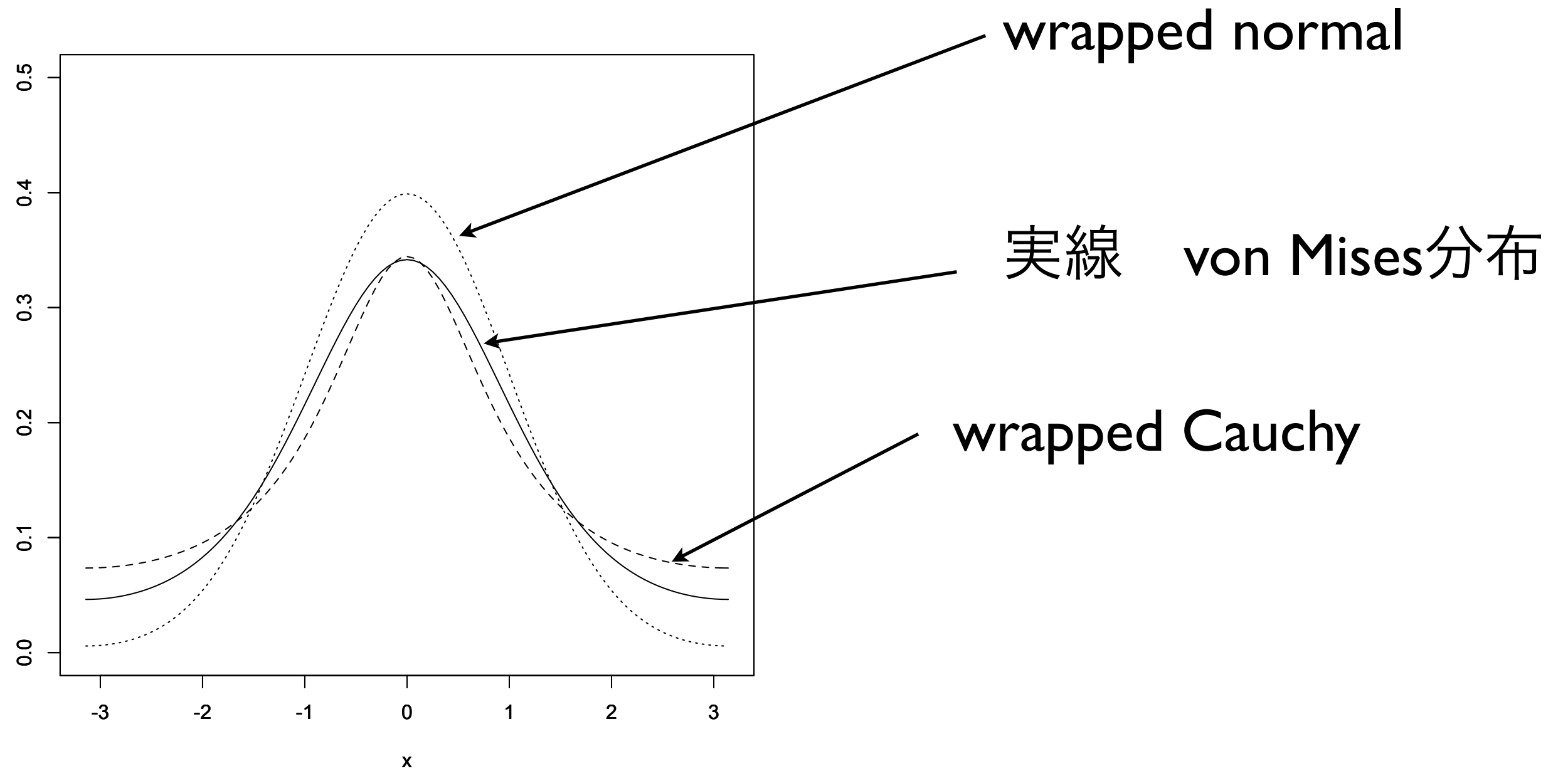
円柱に巻き付けることにより circular な変数の分布



正規分布



# 分布の比較



Rのパッケージ

circular

NPCirc

CircStats

CircNNTSR

isocirc

circular DNA関係のパッケージ

グラフ    Circos

グラフ    RCircos    plotrix



circularなデータは、  
circularなデータ用の方法で分析