パターン認識と機械学習

2.3.8 周期変数

7月**4**日（金）

豊川　幸秀

ガウス分布における問題

- 連続変数の密度モデルとしては適さない場面もある。その例が、周期変数。

- 例として、特定位置における一定期間の風向の値をガウス分布を用いて表すとする。

  →ある一定の方向を原点として定める

２つの観測値θ1=1°、θ2=359°がある。

原点が0°　　:平均180°、標準偏差179°

原点が180°:平均0°、標準偏差1°

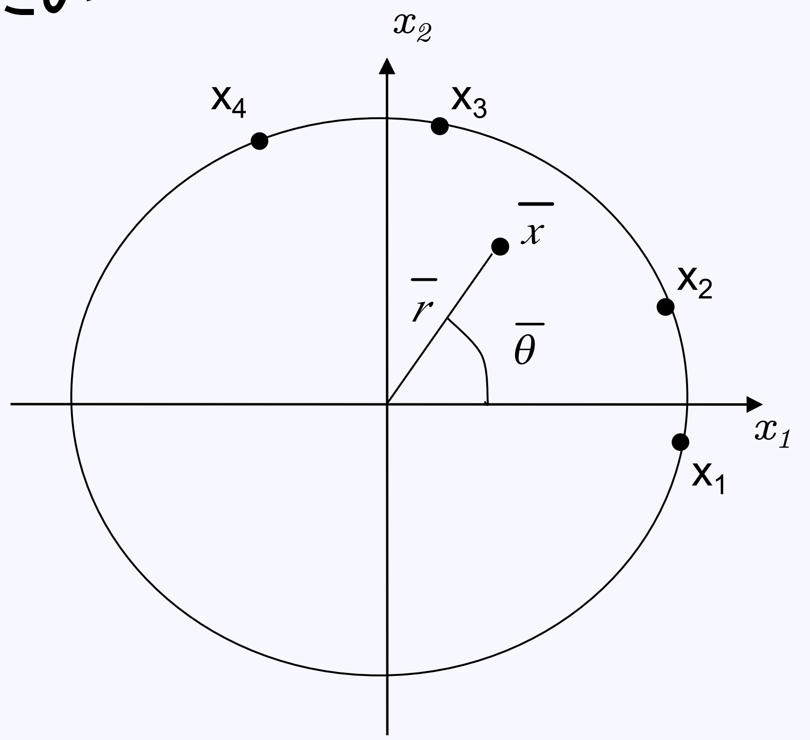
原点に大きく依存してしまうため、
別の方法が必要となる

別のアプローチ

- $D = \{\theta_1, \ldots, \theta_N\}$ の平均を求めたい

- 観測点は、$\| x_n \| = 1, n = 1,\ldots,N$ を満たす**2**次元単位ベクトル $x_1, \ldots, x_N$ で表される。

- $\overline{x} = \frac{1}{N}\sum_{n=1}^{N} x_n$ となる。



各観測値及び単純平均を直交座標で表すと、

- x_n: $x_1 = \cos\theta_n$, $x_2 = \sin\theta_n$
- $\bar{x}$: $\bar{x}_1 = \bar{r}\cos\bar{\theta}$, $\bar{x}_2 = \bar{r}\sin\bar{\theta}$

よって、

$$\bar{r}\cos\bar{\theta} = \frac{1}{N}\sum_{n=1}^{N}\cos\theta_n \qquad \bar{r}\sin\bar{\theta} = \frac{1}{N}\sum_{n=1}^{N}\sin\theta_n$$

すなわち、 $$\bar{\theta} = \tan^{-1}\left\{\frac{\sum_n \sin\theta_n}{\sum_n \cos\theta_n}\right\} \qquad (2.169)$$

フォン･ミーゼス分布

- ガウス分布の周期変数への一般化
- 周期2πの確率密度$p(\theta)$の分布は以下を満たす
  - $p(\theta) \geq 0$
  - $\int_0^{2\pi} p(\theta) d\theta = 1$
  - $p(\theta + M2\pi) = p(\theta)$　M: 任意の整数

これらの条件を満たす２変数 x = (x_1, x_2) 上のガウス分布は、

$$p(x_1, x_2) = \frac{1}{2\pi\sigma^2} \exp\{-\frac{(x_1-\mu_1)^2 + (x_2-\mu_2)^2}{2\sigma^2}\} \quad (2.173)$$

（平均: $\mu = (\mu_1, \mu_2)$ 、共分散行列: $\Sigma = \sigma^2 I$）

極座標に変換

$$x_1 = r\cos\theta, x_2 = r\sin\theta$$

$$\mu_1 = r_0\cos\theta_0, \mu_2 = r_0\sin\theta_0$$

(2.173)に代入して変形。指数部分に注目すると、

$$\frac{r_0}{\sigma^2}\cos(\theta - \theta_0) + const$$ ($const$: θ と独立な項)

$m = \dfrac{r_0}{\sigma^2}$ とおくと、r = 1の単位円において、

$$\boxed{p(\theta \mid \theta_0, m) = \frac{1}{2\pi I_0(m)} \exp\{m\cos(\theta - \theta_0)\}}$$

フォン・ミーゼス分布もしくは循環正規分布と呼ぶ

正規化係数

- $I_0(m)$ で表され、0次の第1種変形ベッセル関数である。

$$I_0(m) = \frac{1}{2\pi}\int_0^{2\pi} \exp\{m\cos\theta\}\,d\theta$$

→mが大きいと、分布が近似的にガウス分布となる

最尤推定量

- フォン・ミーゼス分布の対数尤度関数は

$$\ln p(D|\theta_0, m) = -N\ln(2\pi) - N\ln I_0(m) + m\sum_{n=1}^{N}\cos(\theta_n - \theta_0)$$

θ_0 についての導関数を0とし、最大化すると

$$\theta_0^{ML} = \tan^{-1}\left\{\frac{\sum_n \sin\theta_n}{\sum_n \cos\theta_n}\right\}$$

→(2.169)と一致

同様に、mについても最大化すると、

$$A(m_{ML}) = \frac{1}{N}\sum_{n=1}^{N}\cos(\theta_n - \theta_0^{ML})$$

ただし、$I_0^{'}(m) = I_1(m)$ 、$A(m) = \dfrac{I_1(m)}{I_0(m)}$

その他の手法

- 角座標を一定区間で分割した観測値のヒストグラムを用いる手法　→　制限あり(2.5節)
- ユークリッド空間上のガウス分布から始め、それを周辺化する手法　→　より複雑
- 分布を2πで間隔で、周期変数(0,2π)に写像する手法　→　より複雑
- 一方、フォン・ミーゼス分布は・・・

単峰であることが弱点

→混合分布を用いることにより多峰性を扱える