方向性统计 directional statistics

简介

首先，在圆上一个合适的点切开这个圆，并在最终线上的观测中使用常规的汇总统计，会有一个缺点：太过于依赖切开的那个点。
我们看到，一个圆上大小为2的样本，包括 1° 和 359°。如果从0°切开这个样本，那么样本均值就是180°，样本的标准差就是179°。然而在180°切开这个圆，那么样本均值就是0°，样本的标准差就是1°。

在圆数据上构建一个汇总统计的合适方式，也就是将圆上的点视为平面上的单位向量，并且取向量的样本均值的极坐标。

位置的测量

方向均值

n个单位向量的方向均值θ是$x_1 + x_2 + \ldots + x_n$的方向，也是向量均值 $\bar{x}$的方向，因为二者相差一个标量，方向是一致的。
这个方向均值的笛卡尔坐标如下:

$$(\overline{C}, \overline{S}) = (\frac{1}{n}\sum_{j=1}^{n} cos\theta_j, \frac{1}{n}\sum_{j=1}^{n} sin\theta_j) = (\overline{R}cos\overline{\theta}, \overline{R}sin\overline{\theta})$$

所以这里非常关键的一点就是，圆统计中的 $\overline{\theta}$ 并不是$(\theta_1 + \ldots + \theta_n)/n$, 而是通过$(\overline{C}, \overline{S})$:

$$\overline{\theta} = \begin{cases} \tan^{-1}\left(\frac{\overline{S}}{\overline{C}}\right) & \text{if } C \geq 0 \\ \tan^{-1}\left(\frac{\overline{S}}{\overline{C}} + \pi\right) & \text{if } C < 0 \end{cases}$$

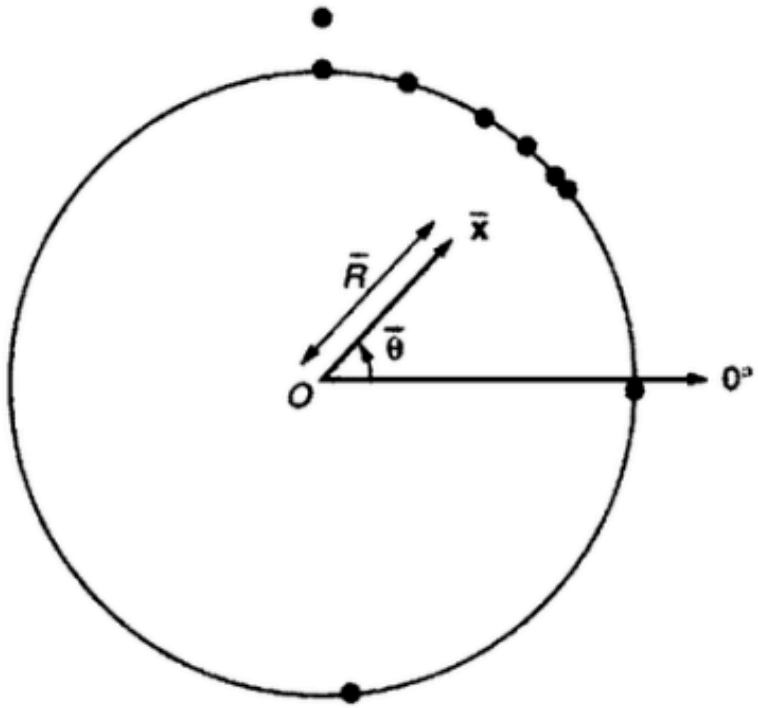

Figure 2.2 The mean direction $\bar{\theta}$ and the mean resultant length $\bar{R}$ for the roulette data in Example 1.1.

集中度和分散度的衡量

均值与圆方差

如果$\theta_1, ..., \theta_n$是簇拥在一起的，那么$\overline{R}$就接近于1，如果$\theta_1, ..., \theta_n$是分散开的的，那么$\overline{R}$就接近于0。所以$\overline{R}$是一个数据集集中程度的标准。当然，如果$\overline{R} \approx 0$ 并不意味着方向在圆上均匀分布。

圆方差：$V = 1 - \overline{R}, 0 \leq V \leq 1$

圆标准差

样本的圆标准差如下定义：

$v = \{-2log(1-V)\}^{1/2} = \{-2log\overline{R}\}^{1/2}$

经过对数处理之后，v的范围就是 $[0, \infty]$，然而V的范围是$[0, 1]$

根据不等式，对于很小的V来说，

$V \approx log(1-V)$

所以可得

$v \approx (2V)^{1/2} = \{2(1-\overline{R})^{1/2}\}$