总体分布的非参数估计

已知样本所属的类别，但未知总体概率密度函数的形式，要求我们直接推断概率密度函数本身。统计学中常见的一些典型分布形式不总是能够拟合实际中的分布。此外，在许多实际问题中经常遇到多峰分布的情况，这就迫使我们必须用样本来推断总体分布，常见的总体类条件概率密度估计方法有 Parzen 窗法和 K 近邻法两种。

非参数估计也有人将其称之为无参密度估计，它是一种对先验知识要求最少，完全依靠训练数据进行估计，而且可以用于任意形状密度估计的方法。常见的非参数估计方法有以下几种：

1. **直方图:** 把数据的值域分为若干相等的区间，数据按照区间分 为若干组，每组形成一个矩形，矩形的高和该组数据的多少成正比， 其底为所属区间，将这些矩形依次排列组成的图形就是直方图。它提 供给数据一个直观的形象，但只适合低维数据的情况，当维数当维数较高时， 直方图所需的空间将随着维数的增加呈指数级增加.

2. **核密度估计**: 原理和直方图有些类似，是一神平滑的无参密度估计方法。对于一组采样数据，把数据的值域分为若干相等的区间， 每个区间称为一个 bin,数据就按区间分为若干组，每组数据的个数和总参数个数的比率就是每个 bin 的概率值。相对于直方图法，它多了一个用于平滑数据的核函数。和密度估计方法适用于中小规模的数据集，可以很快地产生一个渐进无偏的密度估计，有良好的概率统计性质。

3. **K 近邻估计**: 密度估计的加权是以数据点到 x 的欧式距离为基准来进行的，而 K 近邻估计是无论欧式距离多少，只要是离 x 点的最近的 k 个点的其中之一就可以加权.