第九章 概率模型

第 25 讲 非参数估计

黄定江

DaSE @ ECNU
djhuang@dase.ecnu.edu.cn

1. 25.1 直方图估计
2. 25.2 核密度估计
3. 25.3 非参数回归估计
4. 25.4 CDF 和统计泛函的估计

25.1.1 直方图估计的定义

定义 1

直方图可以定义为：

$$\widehat{f}_n(x) = \begin{cases} \widehat{p}_1/h, & x \in B_1 \\ \widehat{p}_2/h, & x \in B_2 \\ \cdots\cdots & \\ \widehat{p}_m/h, & x \in B_m \end{cases}$$

或者可以写的更简洁：

$$\widehat{f}_n(x) = \sum_{j=1}^{m} \frac{\widehat{p}_j}{h} I(x \in B_j)$$

25.1.1 直方图举例

直方图将输入空间划分为 m 个箱子 (bin), 箱子的宽度为 $h = 1/m$ 则这些箱子为 $B_1 = [0, 1/m), B_2 = [1/m, 2/m), \cdots, B_m = [(m-1)/m, 1)$ 计算落入箱子 b 中的样本的数目 V_b, 落入箱子 b 的比率为 $\hat{p}_b = V_b/N$ 则直方图估计为

$$\hat{p}(x) = \sum_{b=1}^{m} \frac{\hat{p}_b}{h} \mathbf{1}(x \in B_b) = \frac{1}{N} \sum_{b=1}^{M} \frac{v_b}{h} \mathbf{1}(x \in B_b)$$

其中 $\mathbf{1}(x \in B_b)$ 表示当 $x \in B_b$ 时其值为 1，否则为 0

25.2.1 核密度估计定义

直方图是不连续的。核密度估计较光滑且比直方图估计较快地收敛到真正的密度。

定义 2

给定一个核 K 与一个正数 h，称作带宽，核密度估计定义为

$$\widehat{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K\left(\frac{x - X_i}{h}\right)$$

其中，参数 h 称为带宽 (bandwidth)，核函数可为任意平滑的函数 K，满足

$$K(u) > 0, \qquad \int K(u)\,du = 1$$

$$\int uK(u)\,du = 0, \qquad \sigma_K^2 = \int u^2 K(u)\,du > 0$$

25.2.1 核密度举例

令 $X1, \cdots, Xn$ 表示观测数据, 它们来自 f 的一个样本. 在本章中, 核定义为任意一个光滑函数 K 使得 $K(x) \geqslant 0$，$\int K(x)\mathrm{d}x = 1$，$\int xK(x)\mathrm{d}x = 0$ 并且 $\sigma_K^2 = \int x^2 K(x)\mathrm{d}x > 0$。核的两个例子分别为 Epanechnikov 核

$$K(x) = \begin{cases} \frac{3}{4}\left(\frac{1-x^2}{5}\right)/\sqrt{5}, & |x| < \sqrt{5} \\ 0, & \text{其他} \end{cases}$$

与高斯（正态）核

$$K(x) = (2\pi)^{-1/2}\mathrm{e}^{-x^2/2}$$

25.2.1 平滑参数 h 的定理

定理 1

在 f 和 K 的弱假设下，

$$R\left(f,\widehat{f}_n\right) \approx \frac{1}{4}\sigma_K^4 h^4 \int \left(f''(x)\right)^2 \,\mathrm{d}x + \frac{\int K^2(x)\mathrm{d}x}{nh}$$

其中，$\sigma_K^2 = \int x^2 K(x)\mathrm{d}x$。最优的带宽为

$$h^* = \frac{c_1^{-2/5} c_2^{1/5} c_3^{-1/5}}{n^{1/5}}$$

其中，$c_1 = \int x^2 K(x)\mathrm{d}x$, $c_2 = \int K(x)^2 \,\mathrm{d}x$ 且 $c_3 = \int \left(f''(x)\right)^2 \,\mathrm{d}x$

证明略。

25.3.1 一元非参数回归

考虑点对 $(x_i, Y_i), \cdots, (x_n, Y_n)$，其关系为

$$Y_i = r(x_i) + \epsilon_i$$

其中，$E(\epsilon_i) = 0$，$r(x_i) = E(Y|X)$。感兴趣的是如何求出 $r(x_i)$。
存在很多非参数回归估计。大多数涉及通过对 Y 取某种加权平均来估计 $r(x)$, 对靠近 x 的点给予更高的权重。一个常用的估计就是所谓 Nadaraya-Watson 核估计。

25.3.1 Nadaraya-Watson 核估计

定义 3

Nadaraya-Watson 核估计定义为

$$\widehat{r}(x) = \sum_{i=1}^{n} w_i(x)\, Y_i$$

其中，K 为一个核且其权重 $w_i(x)$ 由下式给出:

$$w_i(x) = \frac{K\left(\frac{x-x_i}{h}\right)}{\sum_{j=1}^{n} K\left(\frac{x-x_j}{h}\right)}$$

核密度回归还有另一种写法:

$$r(x) = \mathbb{E}(Y \mid X = x) = \int y f(y \mid x)\mathrm{d}y = \frac{\int y f(x,y)\mathrm{d}y}{\int f(x,y)\mathrm{d}y}$$

25.3.1 核估计举例

例 1

图1给出了的宇宙微波背景 (CMB) 数据的拟合情况．该数据包含了 n 对观察值 $(x_1, Y_1), \cdots, (x_n, Y_n)$, 其中, x_i 称作多极矩, Y_i 称作温度变化功率谱估计。所看到的是宇宙微波背景辐射中的声波，这是从宇宙大爆炸中留下来的。若令 $r(z)$ 表示真正的功率谱，则

$$Y_i = r(x_i) + \epsilon_i$$

其中, ϵ_i 是一个均值为 0 的随机误差。$r(z)$ 峰值的位置和大小为了解早期宇宙的状况提供了有价值的线索. 图1给出了基于交叉验证的拟合，既有一个欠光滑的拟合也有一个过光滑的拟合。交叉验证拟合表明了三个定义好的峰值的存在，恰如大爆炸的物理学理论所预测的那样。

Frame Title

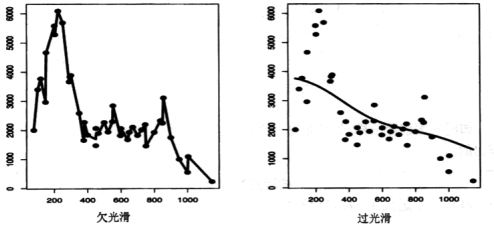

图 1: CMB 数据的回归分析

25.4.1 经验分布函数

令 $X_1, \cdots, X_n \sim F$ 为 IID 样本，其中，F 为实直线上的分布函数，将用经验分布函数估计 F，定义如下：

定义 4

经验分布函数 E 指在每一个数据点 X_i 上的概率密度为 $\frac{1}{n}$ 的 *CDF*，用公式表示为

$$\hat{F}_n(x) = \frac{\sum_{i=1}^{n} I(X_i \leqslant x)}{n}$$

其中，

$$I(X_i \leqslant x) = \begin{cases} 1, & X_i \leqslant x \\ 0, & X_i > x \end{cases}$$

25.4.1 经验分布函数举例

(神经数据) Cox 和 Lewis(1966) 报告了一种神经两次起搏之间的等待时间，共有 799 个数据。图2为经验的 CDF $\hat{F}_n$，数据点以垂直直线体现在图的底部。假设要估计等待时间在 0.4 到 0.6 秒之间的概率，估计值为 $\hat{F}(0.6) - \hat{F}(0.4) = 0.93 - 0.84 = 0.09$。

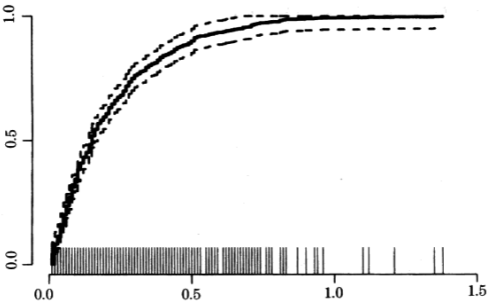

图 2: 神经数据

25.4.1 CDF 定理

定理 2

在任意固定点 x 有

$$\mathbb{E}\left(\hat{F}_n(x)\right) = F(x)$$

$$\mathbb{V}\left(\hat{F}_n(x)\right) = \frac{F(x)(1-F(x))}{n}$$

$$\mathrm{MSE} = \frac{F(x)(1-F(x))}{n} \to 0$$

$$\hat{F}_n(x) \xrightarrow{P} F(x)$$

定理 3

(*Glivenko-Cantelli* 定理) $X_1, \cdots, X_n \sim F$, 则

$$\sup_x \left|\hat{F}_n(x) - F(x)\right| \xrightarrow{p} 0$$

25.4.1 CDF 定理

定理 4

(Dvoretzky-Kiefer-Wolfowitz(DKW) 不等式*)* 令 $X_1, \cdots, X_n \sim F$, 则对任意 $\epsilon > 0$ 有

$$\mathbb{P}\left(\sup_x \left|F(x) - \hat{F}_n(x)\right| > \epsilon\right) \leqslant 2\mathrm{e}^{-2n\epsilon^2}$$

通过 DKW 不等式，可以按如下方式建立置信集：

25.4.1 CDF 定理

定义：

$$L(x) = \max\left\{\hat{F}_n(x) - \epsilon_n, 0\right\}$$
$$U(x) = \min\left\{\hat{F}_n(x) + \epsilon_n, 1\right\}$$

其中，

$$\epsilon = \sqrt{\frac{1}{2n}\log\left(\frac{2}{\alpha}\right)}$$

对任意 F，由4得

$$\mathbb{P}(\text{对所有}x, L(x) \leqslant F(x) \leqslant U(x)) \geqslant 1 - \alpha$$

例 2

图2的虚线给出了 95% 置信带，其中，$\epsilon_n = \sqrt{\frac{1}{2n}\log\left(\frac{2}{0.05}\right)} = 0.048$

统计泛函的举例

统计泛函 $T(F)$ 是 F 的任意函数，例如，均值 $\mu = \int x \, \mathrm{d}F(x)$，方差 $\sigma^2 = \int (x-\mu)^2 \, dF(x)$，中位数 $m = F^{-1}(1/2)$。

定义 5

$\theta = T(F)$ 的的嵌人式估计量定义为

$$\hat{\theta}_n = T\left(\hat{F}_n\right)$$

换言之，就是用经验分布函数 $\hat{F}n$ 代替未知函数 F。

定义 6

如果对函数 $r(x)$ 有 $T(F) = \int r(x) \mathrm{d}F(x)$，则称 T 为线性泛函。

线性泛函的嵌入式估计量

定理 5

线性泛函 $T(F)=\int r(x)\mathrm{d}F(x)$ 的嵌入式估计量为

$$T\left(\hat{F}_n\right)=\int r(x)\mathrm{d}\hat{F}_n(x)=\frac{1}{n}\sum_{i=1}^{n}r\left(X_i\right)$$

例 3

(均值) 令 $\mu=T(F)=\int x\,\mathrm{d}F(x)$，则均值的嵌入式估计量为 $\hat{\mu}=\int x\,\mathrm{d}\hat{F}_n(x)$，标准误差 $\mathrm{se}=\sqrt{\mathbb{V}}\left(\bar{X}_n\right)=\sigma/\sqrt{n}$，如果 $\hat{\sigma}$ 表示 σ 的估计，则估计的标准误差为 $\hat{\sigma}/\sqrt{n}$，的基于正态的置信区间为 $\bar{X}_n\pm z_{\alpha/2}\,\mathrm{se}$.。