似然函数

似然函数

- 给定联合样本值x关于参数θ的函数：$L(\theta|\mathbf{x}) = f(\mathbf{x}|\theta)$
  其中x是随机变量X取得的值，θ是未知的参数。

- $f(\mathbf{x}|\theta)$是密度函数，表示给定θ下的联合密度函数。

- 似然函数是关于θ的函数而密度函数是关于x的函数。

似然函数

- 离散情况下

概率密度函数：$f(\mathbf{x}|\theta) = \mathbb{P}_\theta(\mathbf{X} = \mathbf{x})$

表示在参数θ的下随机变量X取到x的可能性

$$L(\theta_1|\mathbf{x}) = \mathbb{P}_{\theta_1}(\mathbf{X} = \mathbf{x}) > \mathbb{P}_{\theta_2}(\mathbf{X} = \mathbf{x}) = L(\theta_2|\mathbf{x})$$

如果有上式成立，则在参数θ1下随机变量X取到x值的可能性大于θ2

似然函数

连续情况下

如果X是连续随机变量给定足够小的ε>0，那么其在(x-ε,x+ε)内的概率为：

$$\mathbb{P}_\theta(x-\epsilon < X < x+\epsilon) = \int_{x-\epsilon}^{x+\epsilon} f(x|\theta)dx \approx 2\epsilon f(x|\theta) = 2\epsilon L(\theta|x)$$

得到的结果与离散型一致！概率表达了在给定参数θ时X=x的可能性
而似然表示的是在给定样本X=x时，参数的可能性！

极大似然估计

- 谁干掉的多?

- 在一次吃鸡比赛中，有两位选手，一个是职业选手，一个是菜鸟路人。
  比赛结束后，公布结果有一位选手完成20杀，请问是哪个选手呢?

- 估计大家都选的是职业选手吧!
  因为我们会普遍认为概率最大的事件最有可能发生!

- 极大似然估计：在一次抽样中，得到观测值x1,x2...xn。
  选取θ'(x1,x2...xn)作为θ的估计值，使得θ=θ'(x1,x2...xn)时样本出现的概率最大。

极大似然估计

极大似然估计

- 离散型样本：$L(\theta)=\prod_{i=1}^{n}p(x_i;\theta)$
- 连续型样本：$L(\theta)=\prod_{i=1}^{n}f(x_i;\theta)$
- 极大似然估计：$L(x_1,x_2,\cdots,x_n;\hat{\theta})=\max_{\theta\in\Theta}L(x_1,x_2,\cdots,x_n;\theta)$

极大似然估计

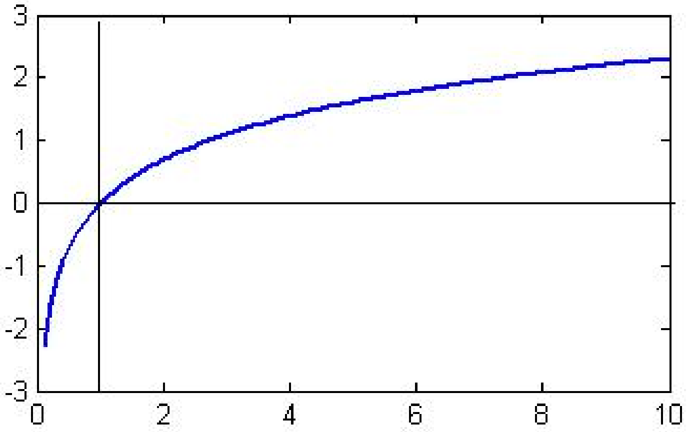
- 极大似然估计求解

  - 构造似然函数：$L(\theta)$

  - 对似然函数取对数：$\ln L(\theta)$

  - 求偏导：$\dfrac{d\ln L}{d\theta}=0$

  - 求解得到θ值



极大似然估计

设 X 服从参数 $\lambda(\lambda>0)$ 的泊松分布，$x_1,x_2,\cdots,x_n$ 是来自于 X 的一个样本值，求 λ 的极大似然估计值.

因为X的分布律为 $P\{X=x\}=\dfrac{\lambda^x}{x!}\mathrm{e}^{-\lambda}$, $(x=0,1,2,\cdots,n)$

所以λ的似然函数为 $L(\lambda)=\prod\limits_{i=1}^{n}\left(\dfrac{\lambda^{x_i}}{x_i\,!}\mathrm{e}^{-\lambda}\right)=\mathrm{e}^{-n\lambda}\dfrac{\lambda^{\sum\limits_{i=1}^{n}x_i}}{\prod\limits_{i=1}^{n}(x_i\,!)}$,

极大似然估计

$$\ln L(\lambda) = -n\lambda + \left(\sum_{i=1}^{n} x_i\right)\ln\lambda - \sum_{i=1}^{n}(x_i!),$$

令 $$\frac{\mathrm{d}}{\mathrm{d}\lambda}\ln L(\lambda) = -n + \frac{\sum_{i=1}^{n} x_i}{\lambda} = 0$$

解得 λ 的极大似然估计值为 $\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x},$

极大似然估计

极大似然估计

离散型样本： $L(\theta)=\prod_{i=1}^{n} p(x_i;\theta)$

连续型样本： $L(\theta)=\prod_{i=1}^{n} f(x_i;\theta)$

极大似然估计： $L(x_1,x_2,\cdots,x_n;\hat{\theta})=\max_{\theta\in\Theta} L(x_1,x_2,\cdots,x_n;\theta)$