Chapter 10 统计推断

△ 关于某参数为 θ 的分布族 $p(x;\theta)$

有数据样本 $(X_1, X_2 \cdots X_n) = \vec{X}$

目标：找出最有可能的 θ（点估计）

2种方法：

- 频率观点：将 θ 视作一个未知数求出来
- 贝叶斯观点：将 θ 视作一个随机变量求出来

最大似然估计

△ Frequentist Perspective (MLE)

流程：未知数 θ $\longrightarrow$ 通过 $p_X(x;\theta)$ 计算 $\longrightarrow$ $\hat{\theta}$（数值）

$p_X(\vec{x};\theta)$：似然函数，表示在参数 θ 下数据样本集 $\vec{x}$ 发生的概率 $\Rightarrow$ 是关于 θ 的函数

$$\therefore \hat{\theta} = \arg\max p_X(\vec{x};\theta) = \arg\max \log p_X(\vec{x};\theta)$$

$\because X_i$ are iid

$$\therefore p_X(\vec{x};\theta) = \prod p_X(x_i;\theta)$$

$$\log p_X(\vec{x};\theta) = \sum \log p_X(x_i;\theta)$$

$$\therefore \hat{\theta} = \arg\max \sum_{i=1}^{n} \log p_X(x_i;\theta)$$

Ex: Biased coin

其中有s个1

取值0.1

一枚 Biased coin，令 Head 概率为 θ，有样本集 $x_1 \cdots x_n$

Sol: $L(\theta) = p_X(\vec{x};\theta) = \prod p_X(x_i;\theta) = \theta^s(1-\theta)^{n-s}$

$$\log L(\theta) = s\log\theta + (n-s)\log(1-\theta)$$

$$\hat{\theta} = \arg\max \log L(\theta) \Rightarrow \begin{cases} (\log L(\theta))' = 0 \\ (\log L(\theta))'' \le 0 \end{cases} \Rightarrow \text{解出即可}$$

△ Bayesian Inference

迭代后的 Θ

流程：Θ（未知随机变量）$\xrightarrow{\text{贝叶斯公式}}$ $\Theta|X$

$\longrightarrow$ 得出 $\Theta|X$ 的分布 PMF 或 PDF

（$p_{\Theta|X}(\theta|X=x)$ 或 $f_{\Theta|X}(\theta|X=x)$）

迭代方法：

	Y discrete	Y continuous
X discrete	$P(Y=y\|X=x) = \frac{P(X=x\|Y=y)P(Y=y)}{P(X=x)}$	$f_Y(y\|X=x) = \frac{P(X=x\|Y=y)f_Y(y)}{P(X=x)}$
X continuous	$P(Y=y\|X=x) = \frac{f_X(x\|Y=y)P(Y=y)}{f_X(x)}$	$f_{Y\|X}(y\|x) = \frac{f_{X\|Y}(x\|y)f_Y(y)}{f_X(x)}$

$$p_{\Theta|X}(\theta|\vec{x}) = \frac{p_\Theta(\theta)\, p_{X|\Theta}(\vec{x}|\theta)}{\sum_{\theta'} p_\Theta(\theta')\, p_{X|\Theta}(\vec{x}|\theta')}$$

$\longrightarrow$ 全概率公式

目标：法一：$\hat{\theta} = E[\Theta|\vec{X}=\vec{x}]$ $\longrightarrow$ 取迭代后分布的期望 $\longrightarrow$ Posterior Mean

法二：$\hat{\theta} = \arg\max_\theta p_{\Theta|X}(\theta|\vec{x})$ $\longrightarrow$ MAP

由贝叶斯公式：取分子：（分母为常数）

$\hat{\theta} = \arg\max_\theta p_\Theta(\theta)\, p_{X|\Theta}(\vec{x}|\theta)$ $\longrightarrow$ 取分布中的概率最大处

△ Beta, Gamma 分布

$X \sim Beta(a,b) \Rightarrow$ $\longrightarrow$ Unif 分布的扩展

$$f(x) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}, \quad 0<x<1$$

特例：$Beta(1,1) = Unif(0,1)$

$$\Gamma(a) = \int_0^\infty x^a e^{-x}\frac{dx}{x} \Rightarrow \begin{cases} \Gamma(a+1) = a\Gamma(a) & (a>0) \\ \Gamma(n) = (n-1)! & (n \in Z^+) \end{cases}$$

$Y \sim Gamma(a, \lambda)$

$$\Rightarrow f(y) = \frac{1}{\Gamma(a)}(\lambda y)^a e^{-\lambda y}\frac{1}{y}, \quad y>0$$

特例：$Gamma(1,\lambda) = Expo(\lambda)$

$Y \sim Gamma(n,\lambda) \Rightarrow Y = \sum_{i=1}^{n} X_i$，$X_i \sim Expo(\lambda)$，$E(Y) = \frac{a}{\lambda}$　$Var(Y) = \frac{a}{\lambda^2}$

若 $\lambda=1 \Rightarrow E(Y) = Var(Y) = n \Rightarrow$ Pois 分布

偷个懒 ⇓

Beta-Gamma Connection

$X+Y$ independent of $\frac{X}{X+Y}$

When we add independent Gamma r.v.s X and Y with the same rate λ, the total $X+Y$ has a Gamma distribution, the fraction $\frac{X}{X+Y}$ has a Beta distribution, and the total is independent of the fraction.

Story: Bank–post Office

While running errands, you need to go to the bank, then to the post office. Let $X \sim \mathrm{Gamma}(a, \lambda)$ be your waiting time in line at the bank, and let $Y \sim \mathrm{Gamma}(b, \lambda)$ be your waiting time in line at the post office (with the same λ for both). Assume X and Y are independent. What is the joint distribution of $T = X + Y$ (your total wait at the bank and post office) and $W = \frac{X}{X+Y}$ (the fraction of your waiting time spent at the bank)?

Tips: 注意 $X+Y$, $\frac{X}{X+Y}$ 独立之前提是

$X、Y \sim Gamma\ (a/b, \lambda)$

△ 共轭先验 (BA estimation)

Introduction: 若先验满足某些分布族，则后验同样满足该分布族，只需改变参数即可

o Beta-Binomial Conjungacy

p 之先验为 $Beta(a,b)$，进行了 n 次试验，k 次成功 $n-k$ 次失败

↳ $f(p) \propto p^{a-1}(1-p)^{b-1}$

$\Rightarrow p$ 之后验为 $Beta(a+k,\ b+n-k)$

$f(p|x=k) \propto p^{a+k-1}(1-p)^{b+n-k-1}$

条件期望取值: $E(p|X=k) = \frac{a+k}{a+b+n}$

若先验为 $Unif(0,1) \Rightarrow \frac{k+1}{n+2}$

再偷个懒 ⇓

Example: Inference of A Biased Coin

1°. $\Theta \sim Unif(0,1) = Beta(1,1)$ # of heads $X|\Theta=\theta \sim Bin(n,\theta)$

By Beta-Binomial Conjugacy, $\Theta | X=k \sim Beta(1+k, 1+n-k)$

$\hat\Theta_{BA} = E[\Theta|X=k] = \frac{1+k}{1+k+1+n-k} = \frac{k+1}{n+2} \Rightarrow \hat\Theta_{BA} = \frac{X+1}{n+2}$.

We wish to estimate the probability of landing heads, denoted by θ, of a biased coin. We model θ as the value of a random variable Θ with a known prior PDF $f_\Theta \sim Unif(0,1)$. We consider n independent tosses and let X be the number of heads observed. Find the MAP estimator of Θ.

posterior PDF

2°. MAP Estimator. $f_{\Theta|X=k}(\theta) \propto \theta^k(1-\theta)^{n-k}$

$\hat\theta_{MAP} = \arg\max_\theta f_{\Theta|X=k}(\theta) = \arg\max_\theta \theta^k(1-\theta)^{n-k}$

$\Rightarrow \hat\theta_{MAP} = \frac{k}{n}$ $\Rightarrow \hat\Theta_{MAP}(X) = \frac{X}{n}$ $\hat\Theta_{MLE} = \frac{X}{n}$

Conclussion: 各种推断/估计方法

MMSE、LLSE (Chapter 8)

↳最小均方误差 ↳线性最小均方误差

MLE(最大似然)、MAP(最大后验概率).

BA(贝叶斯估计) (Chapter 10)

△MMSE:

在给定条件 X 下，求最合适之 Y: $\hat Y_{MMSE} = E(Y|X)$

△LLSE: 线性估计. MMSE之一个特例(将估计函数 $g(x)$ 限定为线性)

△MLE: $p(\vec x;\theta)$ 最大 → 关于 θ 之函数. 解出 $\hat\theta$:

$\hat\theta = \arg\max_\theta p(\vec x;\theta) = \arg\max_\theta \log p(\vec x;\theta)$

△MAP: 由 Θ 过渡出 $\Theta|X$, 得到其分布 $P_{\Theta|X}(\theta|\vec x)$ 或 $f_{\Theta|X}(\theta|\vec x)$

$\hat\theta = \arg\max_\theta P_{\Theta|X}(\theta|\vec x) = \arg\max_\theta P_\Theta(\theta) P_{X|\Theta}(\vec x|\theta)$

△BA: 同族分布改参数