1. 贝叶斯方法综述

频率方法与贝叶斯方法的核心目标都是对参数进行描述，由于两派对待估计参数的理解不同而发展出不同的方法. 贝叶斯学派认为我们可以直接估计 $p(\theta)$, 进而在此基础上“为所欲为”. 这里试举三例：

1. 参数点估计(MAP)：$\theta^* = \underset{\theta}{\operatorname{argmax}}\ p(\theta)$ (Lecture 2)
2. 区间估计： 对 $\forall a, b \in R$，若 $\int_a^b p(\theta)d\theta = 95\%$, 那么区间 $[a, b]$ 即是 θ 的 95%置信区间
3. 结合风险函数进行决策：$\delta^*(D)\underset{a\in A}{argmin} = \int_\Theta L(\theta, a)p(D|\theta)p(\theta)d\theta$ (Lecture 4)
4. 预测：$p(y_*) = \int p(y_*|x_*, w)p(w)dw$ (Lecture 6)

频率学派则认为待估计参数的真实值并非随机变量，故而 $p(\theta)$ 无意义，而应通过“遍历”参数空间来找到我们所要的参数，在参数估计问题中即 $\theta^* = \underset{\theta}{\operatorname{argmax}} SomeCriteron(\theta)$.

以上提及的贝叶斯方法可称之为“狭义”贝叶斯方法. 出于实用性考虑，很多时候人们说的“贝叶斯方法”实际上还包括了 $p(\theta)$ 的估计方法. 在统计学习/机器学习语境下，即是在我们有先验分布 $p(\theta)$ 的基础上，如何根据新的观测数据 (X, Y) 来更新 $p(\theta)$, 亦即 $p(\theta) + (X, Y) \to p(\theta|X, Y)$. 一般而言，只要是使用了 $p(\theta)$ 来解决问题的方法，均可以称之为贝叶斯方法.

贝叶斯方法常使用贝叶斯公式，但不可将二者混为一谈. 在“传统”的参数估计问题中，我们已知随机变量的生成模型 $p(D|\theta)$，并有观测数据 D，希望估计参数 θ. 例如投硬币问题，已知投币 n 次出现 x 次正面的概率 $p(x|n, \theta) = \binom{\mathrm{n}}{\mathrm{x}}\theta^x(1-\theta)^{n-x}$，希望基于 n 和 x 的观测值估计 θ. 在这种问题中我们往往需要用到贝叶斯公式，并结合先验分布，将“生成模型” 转换为“判别模型”：$p(\theta|n, x) = \frac{p(x|n,\theta)p(\theta)}{\int p(x|n,\theta)p(\theta)d\theta}$. 此公式可以有两种理解：若写作 $p(\theta|n, x) = \frac{p(x|n,\theta)}{\int p(x|n,\theta)p(\theta)d\theta}p(\theta)$，可以看作通过“数据更新” $p(\theta) \to \frac{p(x|n,\theta)}{\int p(x|n,\theta)p(\theta)d\theta}p(\theta)$；若写作 $p(\theta|n, x) = \frac{p(\theta)}{\int p(x|n,\theta)p(\theta)d\theta}p(x|n, \theta)$，则可以看作“由果推因” $p(x|n, \theta) \to p(\theta|n, x)$.

从形式上来看，频率方法参数估计如上文所述，通常表现为优化问题，如最大似然或最小损失 . 而贝叶斯方法由于基于先验概率 $p(\theta)$，常表现为概率分布. 例如在线性回归问题中，基于频率方法的最小二乘法的目标是解决优化问题: $\underset{\theta}{\operatorname{argmax}} \sum_i (y_i - w^{\mathrm{T}}x_i)^2$，而噪声服从高斯分布时贝叶斯方法得到的后验概率为 $p(\mathbf{w} \mid D, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \sigma^2) =$

$\mathcal{N}(w\,;\,\boldsymbol{\mu}_{\mathbf{w}|D}, \boldsymbol{\Sigma}_{\mathbf{w}|D})$.

2. 515T 各课概括

- **Lecture 1: Introduction to the Bayesian Method**

  介绍参数估计问题、贝叶斯公式以及贝叶斯方法的一般概念.

- **Lecture 2: Bayesian Inference I (coin flipping)**

  贝叶斯方法在点估计和假设检验中的应用.

- **Lecture 3: Bayesian Inference II (hypothesis testing and summarizing distributions)**

  (待补充)

- **Lecture 4: Bayesian Inference III (decision theory)**

  介绍点估计问题、统计决策问题. (待补充)

- **Lecture 5: The Gaussian Distribution**

  介绍二元高斯分布及其性质，为后续内容作铺垫.

- **Lecture 6: Bayesian Linear Regression**

  高斯噪声下的线性回归. 已在上文提及.

- **Lecture 7: Bayesian Model Selection**

  (待补充)

- **Lecture 8: Bayesian Logistic Regression / The Laplace Approximation**

- **Lecture 9: The Kernel Trick**

- **Lecture 10: Gaussian Process Regression**

- **Lecture 11: Kernels**