BS1 - Introduction

Mou Minghao

SSE, CUHK(SZ)

August 6, 2022

Outline

Introduction

- probability to express our beliefs about unknown quantities
- Bayes' rule provides a rational method for updating beliefs in light of new information

Bayesian Learning

- statistical induction: the process of learning about the general characteristics of a population from a subset of members of that population
- sample space $\mathcal{Y}$; parameter space Θ
- prior distribution $p(\theta)$; sampling model $p(y|\theta)$
- $p(\theta|y) = \frac{p(y|\theta)p(\theta)}{\int_\Theta p(y|\tilde{\theta})p(\tilde{\theta})d\tilde{\theta}}$
- Bayes' rule does not tell what the belief should be but how it should change

Why Bayes?

Recap

Definition

A class $\mathcal{P}$ of prior distributions for θ is called conjugate for a sampling model $p(y|\theta)$ if

$$p(\theta) \in \mathcal{P} \Rightarrow p(\theta|y) \in \mathcal{P} \quad (1)$$

namely, the posterior is invariant under Bayesian update

- **One-parameter exponential family models:** $p(y|\phi) := h(y)c(\phi)e^{(y)}$, ϕ is the unknown parameter and $t(y)$ is the sufficient statistics. Prior distribution of the form $p(\phi|n_0, t_0) := \kappa(n_0, t_0)c(\phi)^{n_0}e^{n_0 t_0 \phi}$
- Recall that if $\theta \sim Beta(a, b)$ and $Y \sim Binomial(n, \theta)$, then

$$\{\theta|Y = y\} \sim Beta(a + y, b + n - y) \quad (2)$$

- Recall that if $\theta \sim Gamma(a, b)$ and $Y_1, ..., Y_n|\theta \sim Poisson(\theta)$, then

$$\{\theta|Y_1, ..., Y_n\} \sim Gamma(a + \sum_{i=1}^{n} Y_i, b + n) \quad (3)$$

The Monte Carlo Method

Monte Carlo approximation, is based on random sampling and its implementation does not require a deep knowledge of calculus or numerical analysis

Suppose we could sample some number S of independent, random θ-values from the posterior distribution $p(\theta|y_1, y_2, ..., y_n)$:

$$\theta^{(1)}, \theta^{(2)}, ..., \theta^{(n)} \sim_{i.i.d.} p(\theta|y_1, y_2, ..., y_n) \tag{4}$$

Then the empirical distribution of the samples would approximate $p(\theta|y_1, y_2, ..., y_n)$, with the approximation improving with increasing S. The empirical distribution of $\left\{\theta^{(1)}, ..., \theta^{(n)}\right\}$ is known as a Monte Carlo approximation to $p(\theta|y_1, ..., y_n)$

The Monte Carlo Method

the law of large numbers implies that

1. $\bar{\theta} = \sum_{s=1}^{S} \theta^{(s)}/S \rightarrow \mathbb{E}[\theta|y_1, ..., y_n]$
2. $\sum_{s=1}^{S}(\theta^{(s)} - \bar{\theta})^2/(S-1) \rightarrow Var[\theta|y_1, ..., y_n]$
3. $\#(\theta^{(s)} \leq c)/S \rightarrow Pr(\theta \leq c|y_1, ..., y_n)$
4. $\cdots$

let $\hat{\sigma}^2 := \sum(\theta^{(s)} - \bar{\theta})^2/(S-1)$ be the Monte Carlo estimate of $Var[\theta|y_1, ..., y_n]$, the Monte Carlo standard error is $\sqrt{\hat{\sigma}^2/S}$. An approximate 95% Monte Carlo confidence interval for the posterior mean of θ is $\hat{\theta} + -2\sqrt{\hat{\sigma}^2/S}$

References I