Bayesian Parameter Estimation

Bayesian Equation

Events

$$S = \{ \text{ simple event that may happen during a random experiment} \}$$
$$X : 2^S \to \mathbb{R}$$

$A, B \subseteq 2^S$

Subset and Equality

$A \subseteq B \iff \text{A is included by B}$

$A \subseteq B \land B \subseteq A \to A = B$

Mutual Exclusion and Negation

If A happens, then B cannot happen, and vice versa.

$\bar{A} = \{A \text{ does not happen}\}$. Exactly one of A and $\bar{A}$ will happen.

$$A \cup \bar{A} = U$$
$$A \cap \bar{A} = \emptyset$$

Addition Law of Probability

When A, B are ME, then

$$A + B \triangleq A \cup B$$
$$AB \triangleq A \cap B$$
$$P(A + B) = P(A) + P(B)$$

Total Probability Equation

Suppose $B_1, \ldots, B_n$ are **mutually exclusive**. $AB_1, \ldots, AB_n$ are mutually exclusive.

Now further suppose that $B_1 + \cdots + B_n = \Omega$. (**collectively exhaustive**)

$$P(A) = P(A\Omega) = P(A(\sum_i B_i)) = P(\sum_i AB_i) = \sum_i P(AB_i)$$

Conditional Probability

$A, B,\ P(B) \neq 0$. $P(A|B)$ is the probability that A happens when B is known to happen.

$$P(A|B) = \frac{P(AB)}{P(B)}$$

Counting measure: M_A the number of simple events that A happens, M_B the number of simple events that B happens, M_{AB} the number of simple events that AB happens.

$$P(A|B) = \frac{M_{AB}}{M_B} = \frac{M_{AB}/M}{M_B/M} = \frac{P(AB)}{P(A)}$$

When $P(A|B) = P(A)$, we say that A and B are **independent**:

$$P(AB) = P(A)P(B)$$

The Bayesian Equation

Suppose $B_1, \ldots, B_n$ are **mutually exclusive and collectively exhaustive**

$$\begin{aligned} P(B_i|A) &= \frac{P(AB_i)}{P(A)} \\ &= \frac{P(A)P(A|B_i)}{P(A)} \\ &= \frac{P(A)P(A|B_i)}{\sum_j P(AB_j)} \\ &= \frac{P(A)P(A|B_i)}{\sum_j P(A)P(A|B_j)} \end{aligned}$$

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{\sum_j P(B_i)P(A|B_j)}$$

$$P(\theta|A) = \frac{P(\theta)P(A|\theta)}{\int P(\theta)P(A|\theta)\mathrm{d}\theta}$$

$$\text{posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{evidence}}$$

Parameter Estimation

$D = (x^1, \ldots, x^n)$ is the dataset.

$$
P(\theta|D) = \frac{P(\theta)P(D|\theta)}{\int P(\theta)P(D|\theta)\mathrm{d}\theta}
$$
$$
\text{where } P(D|\theta) = \prod_i P(x^i|\theta)
$$

The process above mimic the following example.

	Red	**White**	**Black**
1	80	10	10
2	10	80	10
3	10	10	80

$$
A : \text{I draw a red ball from mixed balls}
$$
$$
B_i : \text{The ball I draw is from box } i
$$

From the perspective of conditional probability ,

$$
P(B_1|A) = \frac{P(AB_1)}{P(A)} = \frac{80/300}{100/300}
$$
$$
P(B_2|A) = 10/100
$$
$$
P(B_3|A) = 10/100
$$

From the perspective of Bayesian equation,

$$
P(B_i|A) = \frac{P(B_i)P(A|B_i)}{\sum_j P(B_i)P(A|B_j)}
$$

	1	**2**	**3**
prior ratio	100	100	100
likelihood	80/100	10/100	10/100
posterior ratio	80	10	10
conditional probability	0.8	0.1	0.1

Example: Coin Flip

Suppose in $N = 10$ experiments, there are $n = 4$ times of head up. Denote as the probability of head up as $p \in [0, 1]$.

$$\text{prior: } \rho(p) = U(p; 0, 1) = 1$$
$$\text{likelihood: } P(n|p) = \binom{N}{n} p^n (1-p)^{N-n}$$
$$\text{evidence: } = \int_0^1 \rho(q) P(n|q) \mathrm{d}q = \int_0^1 \binom{N}{n} q^n (1-q)^{N-n} \mathrm{d}q$$
$$\text{posterior} = \rho(p|n, N)$$

$$\begin{aligned}
\rho(p|n, N) &= \frac{\binom{N}{n} p^n (1-p)^{N-n}}{\int_0^1 \binom{N}{n} q^n (1-q)^{N-n} \mathrm{d}q} \\
&= \frac{p^n (1-p)^{N-n}}{\int_0^1 q^n (1-q)^{N-n} \mathrm{d}q} \\
&= \frac{p^n (1-p)^{N-n}}{\mathrm{B}(n+1, N-n+1)} \\
&= \frac{p^{n+1-1} (1-p)^{N-n+1-1}}{\mathrm{B}(n+1, N-n+1)} \\
&= \mathrm{Beta}(p; n+1, N-n+1)
\end{aligned}$$

Point Estimate

Maximum A Posteriori

$$p^* = \arg\max_p \text{posterior}(p)$$

Minimum Mean Squared Error

$$p^* = \arg\min_{\hat{p}} \underset{p \sim \text{posterior}}{\mathrm{E}} (p - \hat{p})^2 \Rightarrow$$
$$p^* = \underset{p \sim \text{posterior}}{\mathrm{E}} p$$

Interval Estimate

Credible Interval

Given a credibility, find the shortest interval such that the probability that the parameter falls in this range is larger than or equal to the credibility.