Sampling

Zhongliang Guo

December 2, 2021

1 Maximum Likelihood (ML)

To estimate the unknown probability.

e.g., an unknown coin

$$P_0 = P(head), \quad 0 < P_0 < 1$$

To estimate P_0, toss it n times,
result in $D = \{d_1, d_2, \cdots, d_n\}$
$i = 1, 2, \cdots, n \qquad d_i \in \{0, 1\}$ (0 for tail, 1 for head).

So,

$$
\begin{aligned}
& P(D|P_0 = p) \\
=& P(d_1|P_0 = p)P(d_2|P_0 = p) \ \cdots \ P(d_n|P_0 = p) \\
=& \prod_{i=1}^{n} P(d_i|p) \qquad\qquad (because\ P(d_i|p)\ under\ Bernoulli\ distribution) \\
=& \prod_{i=1}^{n} p^{d_i}(1-p)^{1-d_i}
\end{aligned}
$$

Take log,

$$
\begin{aligned}
L(p) &= \log P(D|P_0 = p) \\
&= \log \prod_{i=1}^{n} p^{d_i}(1-p)^{1-d_i} \\
&= \sum_{i=1}^{n} \log p^{d_i}(1-p)^{1-d_i} \\
&= \sum_{i=1}^{n} [d_i \log p + (1-d_i)\log(1-p)] \\
&= \sum_{i=1}^{n} d_i \log p + (n - \sum_{i=1}^{n} d_i)\log(1-p)
\end{aligned}
$$

To maximum it, let

$$
\begin{aligned}
\frac{dL}{dp} &= 0 \\
\frac{\sum_{i=1}^{n} d_i}{p} &= \frac{n - \sum_{i=1}^{n} d_i}{1-p} \\
p &= \frac{\sum_{i=1}^{n} d_i}{n}
\end{aligned}
$$

2 Markov chain Monte Carlo (MCMC)

To solve $P(Query|Evidence)$ by sampling

r.v. $X_j \in Query \cup Nuisance$, i.e., $X_j \in Non\!-\!Evidence$

Full conditional distribution:

$$P(X_j|Non\!-\!Evidence/X_j, Evidence)$$

$Non\!-\!Evidence = \{X_1, X_2, \cdots, X_j, \cdots, X_n\}$

2.1 Gibbs Sampling

- Step 0
  Initialise some starting values $x_1^{(0)}, x_2^{(0)}, \cdots, x_n^{(0)}$
  (generally by random guess)

- Step 1
  Repeat
  for $i = 1, \cdots, m$
  for $j = 1, \cdots, n$
  sample $x_j^{(i)} \sim P(X_j|x_1^{(i-1)}, x_2^{(i-1)}, \cdots, x_{j-1}^{(i-1)}, x_{j+1}^{(i-1)}, \cdots, x_n^{(i-1)}, Evidence = e)$
  Keep $[x_1^{(i)}, x_2^{(i)}, \cdots, x_n^{(i)}]$ as a sample

- Step 2
  discard the first a few samples as burn-in (optional) and do Monte Carlo estimation

2.2 Full conditional simplification

$$P(X_j|Non\!-\!Evidence/X_j, Evidence = e) = P(X_j|markov\_blanket(X_j))$$

$$p.s.,\ \ markov\_blanket(X_j)\ \ includes \begin{cases} parents of X_j \\ children of X_j \\ co-parents of the Children of X_j \end{cases}$$

Then $P(X_j|markov\_blanket(X_j))$ can be denoted as the product form as probability in CPTs of Beyesian Network.