

Machine Learnig Homework 4

1 Problem 1

Let $|H|$ denote number of heads and $|T|$ number of tails. We are interested in sequences like: 'H', 'TH', 'TTTH' etc. Since these are coin tosses, probability of events corresponding to these sequences are $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$ and so on. Therefore:

$$E[|H|] = \sum_{i=1}^{\infty} \frac{1}{2^i} = 1 \quad (1)$$

$$E[|T|] = 0 + \frac{1}{4} \cdot 1 + \frac{1}{8} \cdot 2 + \frac{1}{16} \cdot 3 + \dots = \frac{1}{2} \sum_{i=1}^{\infty} \frac{1}{2^i} \cdot i = \frac{1}{2} \cdot 2 = 1 \quad (2)$$

2 Problem 2

Let U number of urns, M number of balls in each urn, N number of balls draw with replacement, n_B number of back balls draw, B the event of drawing a single black ball. We know that the probability of drawing a single black ball from the urn u is $P(B|u) = \frac{u}{M}$ and the probability of drawing n_B black urns from the same urn is $P(n_B|u) = \binom{N}{n_B} P(B|u)^{n_B} (1 - P(B|u))^{N-n_B}$. Due to the Bayes rule:

$$P(u|n_B) = \frac{P(n_B|u)P(u)}{P(n_B)} \quad (3)$$

where $P(u) = \frac{1}{U}$ and

$$P(n_B) = \sum_{u=1}^U P(n_B|u)P(u) = \frac{1}{U} \sum_{u=1}^U \binom{N}{n_B} \left(\frac{u}{M}\right)^{n_B} \left(1 - \frac{u}{M}\right)^{N-n_B} \quad (4)$$

Since we don't know which urn we have, we must rely on the data for the purpose of computing the probability of the next drawn ball being black. It is the MLE estimate of $\theta = P(B|u) = \frac{n_B}{M} = 0.3$

3 Problem 3

$$P(X = x|\theta) = \theta^x (1 - \theta)^{1-x} \quad (5)$$

$$\log(P(X = x|\theta)) = \log(\theta^x (1 - \theta)^{1-x}) = x \log \theta + (1 - x) \log(1 - \theta) \quad (6)$$

$$\theta_{MLE} \Leftrightarrow 0 = \frac{\partial}{\partial \theta} \log(P(X = x|\theta)) = \frac{x}{\theta} - \frac{1-x}{1-\theta} \quad (7)$$

$$\theta_{MLE} = x \quad (8)$$

4 Problem 4

We have that the prior is given by $Beta(\mu, a, b)$ with $E[Beta(\mu, a, b)] = \frac{a}{a+b}$, the posterior is given by $Beta(\mu, m + a, N - m + b)$ with $E[Beta(\mu, m + a, N - m + b)] = \frac{m+a}{N+a+b}$ and $\mu_{MLE} = \frac{m}{N}$.

$$\begin{aligned}
 E[Beta(\mu, m + a, N - m + b)] &= \\
 &= \frac{m + a}{N + a + b} \\
 &= \frac{a + b}{N + a + b} \frac{m + a}{a + b} \\
 &= \frac{a + b}{N + a + b} \frac{a}{a + b} + \frac{N}{N + a + b} \frac{m}{N} \\
 &= \lambda \frac{a}{a + b} + (1 - \lambda) \frac{m}{N} \\
 &= \lambda E[Beta(\mu, a, b)] + (1 - \lambda) \mu_{MLE}
 \end{aligned} \tag{9}$$

5 Problem 5

For n i.i.d. samples $x_i \sim Poi(\lambda) = \frac{e^{-\lambda} \lambda^x}{x!}$ we have

$$P(x_1, \dots, x_n | \lambda) = \prod_{i=1}^n P(x_i | \lambda) = \frac{e^{-n\lambda} \lambda^{\sum_{i=1}^n x_i}}{\sum_{i=1}^n x_i!} \tag{10}$$

$$\log(P(x_1, \dots, x_n | \lambda)) \propto -n\lambda + \log(\lambda) \sum_{i=1}^n x_i \tag{11}$$

$$\frac{\partial}{\partial \lambda} \log(P(x_1, \dots, x_n | \lambda)) \propto -n + \frac{1}{\lambda} \sum_{i=1}^n x_i = 0 \tag{12}$$

$$\lambda_{MLE} = \frac{1}{n} \sum_{i=1}^n x_i \tag{13}$$

It is unbiased, since due to the linearity of the expected value operator we have:

$$E[\lambda_{MLE}] = E\left[\frac{1}{n} \sum_{i=1}^n x_i\right] = \frac{1}{n} \sum_{i=1}^n E[x_i] = \frac{1}{n} \sum_{i=1}^n \lambda = \lambda \tag{14}$$

Gamma distribution is given by

$$Gamma(\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} \propto \lambda^{\alpha-1} e^{-\beta\lambda} \tag{15}$$

and the posterior:

$$\begin{aligned}
& P(\lambda|\mathbf{D}, \alpha, \beta) \\
& \propto P(\mathbf{D}|\lambda)P(\lambda|\alpha, \beta) \\
& = e^{-n\lambda} \lambda^{\sum_{i=1}^n x_i} \lambda^{\alpha-1} e^{-\beta\lambda} \\
& = \lambda^{\alpha-1+\sum_{i=1}^n x_i} e^{-(n+\beta)\lambda}
\end{aligned} \tag{16}$$

therefore

$$\frac{\partial}{\partial \lambda} \log(P(\lambda|\mathbf{D}, \alpha, \beta)) \propto -n + \beta + \frac{1}{\lambda} \left(\alpha - 1 + \sum_{i=1}^n x_i \right) = 0 \tag{17}$$

finally

$$\lambda_{MAP} = \frac{1}{n + \beta} \left(\alpha - 1 + \sum_{i=1}^n x_i \right) \tag{18}$$