EE5907/EE5027 Week 2: Probabilistic Estimation + Conjugate Priors

Exercise 3.1

The likelihood is given by

$$p(\mathcal{D}|\theta) = \theta^{N_1}(1-\theta)^{N_0} \tag{1}$$

Hence the log-likelihood is given by

$$\log p(\mathcal{D}|\theta) = N_1 \log\theta + N_0 \log(1-\theta) \tag{2}$$

To optimize the log-likelihood, we get

$$\underset{\theta}{\operatorname{argmax}}\, p(\mathcal{D}|\theta) = \underset{\theta}{\operatorname{argmax}}(N_1 \log\theta + N_0 \log(1-\theta)) \tag{3}$$

Differentiating with respect to θ and set to 0, we get:

$$\begin{aligned}
&\frac{N_1}{\theta} - \frac{N_0}{1-\theta} = 0 \\
&\implies N_1(1-\theta) = N_0\theta \\
&\implies \theta = \frac{N_1}{N_1+N_0} \\
&\implies \theta = \frac{N_1}{N}
\end{aligned}$$

Hence, $\hat{\theta}_{MLE} = \frac{N_1}{N}$

Exercise 3.6

The Poisson distribution can be represented as:

$$\mathcal{D} = (x_1, x_2, \cdots, x_n), \mathcal{D} \sim Poi(\lambda) \tag{4}$$

The likelihood is given by

$$p(\mathcal{D}|\lambda) = \prod_{i=1}^{n} e^{-\lambda}\frac{\lambda^{x_i}}{x_i!} \tag{5}$$

To optimize the log-likelihood, we get

$$
\begin{aligned}
\hat{\lambda}_{MLE} &\triangleq \underset{\lambda}{\operatorname{argmax}} \log \prod_{i=1}^{n} e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} \\
&= \underset{\lambda}{\operatorname{argmax}} \sum_{i=1}^{n} \log \left(e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} \right) \\
&= \underset{\lambda}{\operatorname{argmax}} \sum_{i=1}^{n} \left(-\lambda + x_i \log \lambda - \log x_i! \right) \\
&= \underset{\lambda}{\operatorname{argmax}} \left(-n\lambda + \sum_{i=1}^{n} x_i \log \lambda - \sum_{i=1}^{n} \log x_i! \right) \\
&= \underset{\lambda}{\operatorname{argmax}} \left(-n\lambda + \sum_{i=1}^{n} x_i \log \lambda \right)
\end{aligned}
$$

Differentiating with respect to λ and set to 0, we get:

$$
\begin{aligned}
&-n + \frac{1}{\lambda} \sum_{i=1}^{n} x_i = 0 \\
\Longrightarrow\ &\hat{\lambda}_{MLE} = \frac{1}{n} \sum_{i=1}^{n} x_i
\end{aligned}
$$

Exercise 3.7

a. Multiply the likelihood by the conjugate prior given in the question, we get the following posterior:

$$
\begin{aligned}
& p(\lambda|\mathcal{D}) \propto p(\mathcal{D}|\lambda) p(\lambda) \propto e^{-n\lambda} \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!} \lambda^{a-1} e^{-\lambda b} \\
\Longrightarrow\ & p(\lambda|\mathcal{D}) \propto \frac{1}{\prod_{i=1}^{n} x_i!} e^{-(n+b)\lambda} \lambda^{a-1+\sum_{i=1}^{n} x_i} \\
\Longrightarrow\ & p(\lambda|\mathcal{D}) \propto \lambda^{a-1+\sum_{i=1}^{n} x_i} e^{-(n+b)\lambda} \\
\Longrightarrow\ & p(\lambda|\mathcal{D}) = Ga\left(\lambda | \, a + \sum_{i=1}^{n} x_i, n+b \right)
\end{aligned}
$$

b. Given the mean of Gamma distribution $Ga\,(a, b)$ is $\frac{a}{b}$, we can get the mean of $p(\lambda|\mathcal{D})$ to be

$$
\bar{\theta} = \frac{a + \sum_{i=1}^{n} x_i}{n+b} \tag{6}
$$

Given that $a \to 0$ and $b \to 0$, we have

$$\lim_{a\to 0, b\to 0} \frac{a + \sum_{i=1}^{n} x_i}{n+b} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}$$

Hence, the posterior mean converges to the ML solution.

Exercise 3.12

a. The posterior of the Bernoulli

$$p(\theta|\mathcal{D}) \propto p(\mathcal{D}|\theta)p(\theta)$$

if $\theta = 0.5$,

$$\begin{aligned} & p(\mathcal{D}|\theta)p(\theta) = 0.5^{N+1} \\ \implies & \log p(\mathcal{D}|\theta)p(\theta) = (N+1)\log 0.5 \end{aligned}$$

if $\theta = 0.4$,

$$\begin{aligned} & p(\mathcal{D}|\theta)p(\theta) = 0.4^{N_1} 0.6^{N-N_1} 0.5 \\ \implies & \log p(\mathcal{D}|\theta)p(\theta) = N_1 \log 0.4 + (N - N_1)\log 0.6 + \log 0.5 \end{aligned}$$

if $\theta =$others,

$$p(\mathcal{D}|\theta)p(\theta) = 0$$

For 0.5 to win out over 0.4,

$$\begin{aligned} & (N+1)\log 0.5 > N_1 \log 0.4 + (N - N_1)\log 0.6 + \log 0.5 \\ \implies & N \log \frac{0.5}{0.6} > N_1 \log \frac{0.4}{0.6} \\ \implies & \frac{N_1}{N} > \frac{\log 5/6}{\log 2/3} = \frac{\log 1.2}{\log 1.5} = 0.4497 \text{ because } \log 2/3 \text{ is negative} \end{aligned}$$

Therefore, we have

$$\hat{\theta}_{MAP} = \begin{cases} 0.4 & \text{if } \frac{N_1}{N} < \frac{\log 1.2}{\log 1.5} \\ 0.5 & \text{if } \frac{N_1}{N} > \frac{\log 1.2}{\log 1.5} \end{cases}$$

Note that N_1/N can never be exactly equal to $\frac{\log 1.2}{\log 1.5}$ because $\frac{\log 1.2}{\log 1.5}$ is irrational.

b. If N is large, then the MAP estimate (with the usual beta prior) will approach the true value of 0.41. However, the biased-coin prior will still lead to an estimate of 0.4, resulting in a difference of 0.01 from the true value. Therefore the biased-coin prior does not lead to a consistent estimator.

   If N is small, the unbiased coin prior might possibly be off by a lot. For example, if $N = 1$ and the outcome of the coin toss is head. Then if $\alpha = \beta = 1$, the unbiased coin prior would lead to a MAP estimate of $\hat{\theta} = 1$. On the other hand, the biased coin prior will lead to a MAP estimate of 0.5, which is not that different from 0.41.

Exercise 3.14

a. Denote the counts of each alphabet by N_j. If we use a $\mathrm{Dir}(\alpha)$ prior for θ, the posterior predictive is just

$$p(x = k|\mathcal{D}) = \frac{\alpha_k + N_k}{\sum_{k'}(\alpha_{k'} + N_{k'})}$$

   Substitute $a_k = 10$ and the number of "e" is 260, we have

$$\begin{aligned} p(x_{2001} = \mathrm{e}|\mathcal{D}) &= \frac{10 + 260}{270 + 2000} \\ &= 0.119 \end{aligned}$$

b. Similar to part (a), we easily derive that

$$\begin{aligned} p(x_{2001} = \mathrm{p}|\mathcal{D}) &= \frac{10 + 87}{270 + 2000} \\ &= 0.043 \end{aligned}$$

   Then we have,

$$\begin{aligned} p\left(x_{2001} = p, x_{2002} = a \middle| D\right) &= P\left(x_{2002} = a \middle| x_{2001} = p, D\right) P\left(x_{2001} = P \middle| D\right) \\ &= \frac{\alpha_j + N_j}{\sum_{j'}\left(\alpha_{j'} + N_{j'}\right)} P\left(x_{2001} = P|D\right) \\ &= \frac{10 + 100}{270 + 2001} * 0.043 \\ &= 0.00207 \end{aligned}$$