Exercise 3.1 MLE for the Bernoulli/ binomial model

$$\begin{aligned}
\frac{d}{d\theta}p(D|\theta) &= \frac{d}{d\theta}(\theta^{N_1}(1-\theta)^{N_0}) && (1)\\
&= N_1\theta^{N_1-1}(1-\theta)^{N_0} - N_0\theta^{N_1}(1-\theta)^{N_0 1} && (2)\\
&= \theta^{N_1-1}(1-\theta)^{N_0-1}(N_1(1-\theta) - N_0\theta) && (3)\\
&= \theta^{N_1-1}(1-\theta)^{N_0-1}(N_1 - N\theta) && (4)\\
\therefore \theta_{\text{MLE}} &= \frac{N_1}{N} && (5)
\end{aligned}$$

Exercise 3.2 Marginal likelihood for the Beta-Bernoulli model

$$\begin{aligned}
p(D) &= \frac{[(\alpha_1)\cdots(\alpha_1+N_1-1)][(\alpha_0)\cdots(\alpha_0+N_0-1)]}{(\alpha)\cdots(\alpha+N-1)} && (6)\\
&= \frac{[(\alpha_1)\cdots(\alpha_1+N_1-1)][(\alpha_0)\cdots(\alpha_0+N_0-1)]}{(\alpha_1+\alpha_0)\cdots(\alpha_1+\alpha_0+N-1)} && (7)\\
&= \frac{\Gamma(\alpha_1+\alpha_0)}{\Gamma(\alpha_1+\alpha_0+N)}[(\alpha_1)\cdots(\alpha_1+N_1-1)][(\alpha_0)\cdots(\alpha_0+N_0-1)] && (8)\\
&= \frac{\Gamma(\alpha_1+\alpha_0)}{\Gamma(\alpha_1+\alpha_0+N)}\frac{\Gamma(\alpha_1+N_1)}{\Gamma(\alpha_1)}\frac{\Gamma(\alpha_0+N_0)}{\Gamma(\alpha_0)} && (9)\\
&= \frac{\Gamma(\alpha_1+N_1)\Gamma(\alpha_0+N_0)}{\Gamma(\alpha_1+\alpha_0+N)}\frac{\Gamma(\alpha_1+\alpha_0)}{\Gamma(\alpha_1)\Gamma(\alpha_0)} && (10)
\end{aligned}$$

Exercise 3.3 Posterior prdictive for Beta-Binomial model

$$\begin{aligned}
Bb(1|\alpha_1',\alpha_0',1) &= \frac{B(1+\alpha_1', 1-1+\alpha_0')}{B(\alpha_1',\alpha_0')}\binom{1}{1} && (11)\\
&= \frac{B(1+\alpha_1',\alpha_0')}{B(\alpha_1',\alpha_0')} && (12)\\
&= \frac{\Gamma(\alpha_1'+\alpha_0')}{\Gamma(\alpha_1')\Gamma(\alpha_0')}\frac{\Gamma(1+\alpha_1')\Gamma(\alpha_0')}{\Gamma(1+\alpha_1'+\alpha_0')} && (13)\\
&= \frac{\Gamma(\alpha_1'+\alpha_0')}{\Gamma(\alpha_1')\Gamma(\alpha_0')}\frac{\alpha_1'\Gamma(\alpha_1')\Gamma(\alpha_0')}{(\alpha_1'+\alpha_0')\Gamma(\alpha_1'+\alpha_0')} && (14)\\
&= \frac{\alpha_1'}{\alpha_1'+\alpha_0'} && (15)
\end{aligned}$$

Exercise 3.4 Beta updating from censored likelihood

$$\begin{aligned}
p(\theta, X < 3) &= p(\theta)p(X < 3|\theta) && (16)\\
&= p(\theta)(\sum_{k=0}^{2} p(X = k|\theta)) && (17)\\
&= p(\theta)(\sum_{k=0}^{2} \theta^k (1-\theta)^{(5-k)}) && (18)\\
&= Beta(\theta|1,1)(\sum_{k=0}^{2} \theta^k (1-\theta)^{(5-k)}) && (19)\\
&= \sum_{k=0}^{2} \theta^k (1-\theta)^{(5-k)} && (20)
\end{aligned}$$

Exercise 3.5 Uninformative prior for log-odds ratio

$$\begin{aligned}
p(\theta) &= p(\phi)|\frac{d\phi}{d\theta}| && (21)\\
&= p(\phi)\theta^{-1}(1-\theta)^{-1} && (22)\\
&\propto Beta(\theta|0,0) \ (\because p(\phi) \propto 1) && (23)
\end{aligned}$$

Exercise 3.6 MLE for the Poisson distribution

$$\begin{aligned}
D &= \{x_1, x_2, \dots, x_N\} && (24)\\
p(D|\lambda) &= \Pi_{i=1}^{N} Poi(x_i|\lambda) && (25)\\
&= e^{-N\lambda} \frac{\lambda^{\sum_{i=1}^{N} x_i}}{\Pi_{i=1}^{N}(x_i!)} && (26)\\
\log p(D|\lambda) &= -N\lambda + \sum_{i=1}^{N} x_i \log\lambda - \sum_{i=1}^{N} \log x_i! && (27)\\
\frac{\partial}{\partial\lambda} \log p(D|\lambda) &= -N + \frac{1}{\lambda}\sum_{i=1}^{N} x_i && (28)\\
\therefore \lambda_{MLE} &= \frac{1}{N}\sum_{i=1}^{N} x_i && (29)
\end{aligned}$$

Exercise 3.7 Bayesian analysis of the Poisson distribution

(a)

$$\begin{aligned}
p(\lambda|D) &\propto p(\lambda)p(D|\lambda) && (30)\\
&\propto \lambda^{a-1}e^{-\lambda b}e^{-N\lambda}\frac{\lambda^{\sum_{i=1}^{N}x_i}}{\Pi_{i=1}^{N}(x_i!)} && (31)\\
&= e^{-\lambda(N+b)}\frac{\lambda^{\sum_{i=1}^{N}x_i}}{\Pi_{i=1}^{N}(x_i!)} && (32)\\
&\propto Ga(\lambda|a+\sum_{i=1}^{N}x_i, b+N) && (33)
\end{aligned}$$

(b)

$$\begin{aligned}
\frac{a+\sum_{i=1}^{N}x_i}{b+N} &\to \frac{1}{N}\sum_{i=1}^{N}x_i && (34)\\
&= \lambda_{MLE} && (35)
\end{aligned}$$

Exercise 3.8 MLE for the uniform distribution

(a)

$$\begin{aligned}
D &= \{x_1,\dots,x_N\} && (36)\\
p(D|a) &= \Pi_{i=1}^{N}\frac{1}{2a}I(x_i\in[-a,a]) && (37)
\end{aligned}$$

If $\forall i\ \ -a \le x_i \le a$, then $p(D|a) = \frac{1}{(2a)^n}$. More smaller a, more larger $p(D|a)$.

$\hat{a} = \max\{|x_1|,\dots,|x_N|\}$

(b)

$$\begin{aligned}
p(x_{n+1}|\hat{a}) &= \frac{1}{2\hat{a}}I(x_{n+1}\in[-\hat{a},\hat{a}]) && (38)\\
&= \begin{cases} 0 & (x_{n+1}\notin[-\hat{a},\hat{a}]) \\ \frac{1}{2\hat{a}} & (x_{n+1}\in[-\hat{a},\hat{a}]) \end{cases} && (39)
\end{aligned}$$

(c)

If we use MLE approach, the probability between $-\hat{a}$ and $\hat{a}$ is 0. Bayesian approach is better and introduce a wide range prior.