

Exercise 3.1 MLE for the Bernoulli/ binomial model

$$\frac{d}{d\theta} p(D|\theta) = \frac{d}{d\theta} (\theta^{N_1} (1-\theta)^{N_0}) \quad (1)$$

$$= N_1 \theta^{N_1-1} (1-\theta)^{N_0} - N_0 \theta^{N_1} (1-\theta)^{N_0-1} \quad (2)$$

$$= \theta^{N_1-1} (1-\theta)^{N_0-1} (N_1 (1-\theta) - N_0 \theta) \quad (3)$$

$$= \theta^{N_1-1} (1-\theta)^{N_0-1} (N_1 - N\theta) \quad (4)$$

$$\therefore \theta_{\text{MLE}} = \frac{N_1}{N} \quad (5)$$

Exercise 3.2 Marginal likelihood for the Beta-Bernoulli model

$$p(D) = \frac{[(\alpha_1) \cdots (\alpha_1 + N_1 - 1)][(\alpha_0) \cdots (\alpha_0 + N_0 - 1)]}{(\alpha) \cdots (\alpha + N - 1)} \quad (6)$$

$$= \frac{[(\alpha_1) \cdots (\alpha_1 + N_1 - 1)][(\alpha_0) \cdots (\alpha_0 + N_0 - 1)]}{(\alpha_1 + \alpha_0) \cdots (\alpha_1 + \alpha_0 + N - 1)} \quad (7)$$

$$= \frac{\Gamma(\alpha_1 + \alpha_0)}{\Gamma(\alpha_1 + \alpha_0 + N)} [(\alpha_1) \cdots (\alpha_1 + N_1 - 1)][(\alpha_0) \cdots (\alpha_0 + N_0 - 1)] \quad (8)$$

$$= \frac{\Gamma(\alpha_1 + \alpha_0)}{\Gamma(\alpha_1 + \alpha_0 + N)} \frac{\Gamma(\alpha_1 + N_1)}{\Gamma(\alpha_1)} \frac{\Gamma(\alpha_0 + N_0)}{\Gamma(\alpha_0)} \quad (9)$$

$$= \frac{\Gamma(\alpha_1 + N_1) \Gamma(\alpha_0 + N_0)}{\Gamma(\alpha_1 + \alpha_0 + N)} \frac{\Gamma(\alpha_1 + \alpha_0)}{\Gamma(\alpha_1) \Gamma(\alpha_0)} \quad (10)$$

Exercise 3.3 Posterior predictive for Beta-Binomial model

$$Bb(1|\alpha'_1, \alpha'_0, 1) = \frac{B(1 + \alpha'_1, 1 - 1 + \alpha'_0)}{B(\alpha'_1, \alpha'_0)} \binom{1}{1} \quad (11)$$

$$= \frac{B(1 + \alpha'_1, \alpha'_0)}{B(\alpha'_1, \alpha'_0)} \quad (12)$$

$$= \frac{\Gamma(\alpha'_1 + \alpha'_0)}{\Gamma(\alpha'_1) \Gamma(\alpha'_0)} \frac{\Gamma(1 + \alpha'_1) \Gamma(\alpha'_0)}{\Gamma(1 + \alpha'_1 + \alpha'_0)} \quad (13)$$

$$= \frac{\Gamma(\alpha'_1 + \alpha'_0)}{\Gamma(\alpha'_1) \Gamma(\alpha'_0)} \frac{\alpha'_1 \Gamma(\alpha'_1) \Gamma(\alpha'_0)}{(\alpha'_1 + \alpha'_0) \Gamma(\alpha'_1 + \alpha'_0)} \quad (14)$$

$$= \frac{\alpha'_1}{\alpha'_1 + \alpha'_0} \quad (15)$$

Exercise 3.4 Beta updating from censored likelihood

$$p(\theta, X < 3) = p(\theta)p(X < 3|\theta) \quad (16)$$

$$= p(\theta)\left(\sum_{k=0}^2 p(X = k|\theta)\right) \quad (17)$$

$$= p(\theta)\left(\sum_{k=0}^2 \theta^k (1 - \theta)^{(5-k)}\right) \quad (18)$$

$$= \text{Beta}(\theta|1, 1)\left(\sum_{k=0}^2 \theta^k (1 - \theta)^{(5-k)}\right) \quad (19)$$

$$= \sum_{k=0}^2 \theta^k (1 - \theta)^{(5-k)} \quad (20)$$

Exercise 3.5 Uninformative prior for log-odds ratio

$$p(\theta) = p(\phi) \left| \frac{d\phi}{d\theta} \right| \quad (21)$$

$$= p(\phi) \theta^{-1} (1 - \theta)^{-1} \quad (22)$$

$$\propto \text{Beta}(\theta|0, 0) \quad (\because p(\phi) \propto 1) \quad (23)$$

Exercise 3.6 MLE for the Poisson distribution

$$D = \{x_1, x_2, \dots, x_N\} \quad (24)$$

$$p(D|\lambda) = \prod_{i=1}^N \text{Poi}(x_i|\lambda) \quad (25)$$

$$= e^{-N\lambda} \frac{\lambda^{\sum_{i=1}^N x_i}}{\prod_{i=1}^N (x_i!)} \quad (26)$$

$$\log p(D|\lambda) = -N\lambda + \sum_{i=1}^N x_i \log \lambda - \sum_{i=1}^N \log x_i! \quad (27)$$

$$\frac{\partial}{\partial \lambda} \log p(D|\lambda) = -N + \frac{1}{\lambda} \sum_{i=1}^N x_i \quad (28)$$

$$\therefore \lambda_{MLE} = \frac{1}{N} \sum_{i=1}^N x_i \quad (29)$$

Exercise 3.7 Bayesian analysis of the Poisson distribution

(a)

$$p(\lambda|D) \propto p(\lambda)p(D|\lambda) \quad (30)$$

$$\propto \lambda^{a-1} e^{-\lambda b} e^{-N\lambda} \frac{\lambda^{\sum_{i=1}^N x_i}}{\prod_{i=1}^N (x_i!)} \quad (31)$$

$$= e^{-\lambda(N+b)} \frac{\lambda^{\sum_{i=1}^N x_i}}{\prod_{i=1}^N (x_i!)} \quad (32)$$

$$\propto Ga(\lambda|a + \sum_{i=1}^N x_i, b + N) \quad (33)$$

(b)

$$\frac{a + \sum_{i=1}^N x_i}{b + N} \rightarrow \frac{1}{N} \sum_{i=1}^N x_i \quad (34)$$

$$= \lambda_{MLE} \quad (35)$$

Exercise 3.8 MLE for the uniform distribution

(a)

$$D = \{x_1, \dots, x_N\} \quad (36)$$

$$p(D|a) = \prod_{i=1}^N \frac{1}{2a} I(x_i \in [-a, a]) \quad (37)$$

If $\forall i \quad -a \leq x_i \leq a$, 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)

$$p(x_{n+1}|\hat{a}) = \frac{1}{2\hat{a}} I(x_{n+1} \in [-\hat{a}, \hat{a}]) \quad (38)$$

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

(c)

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

Exercise 3.9 Bayesian analysis of the uniform distribution

$$p(\theta|D) = \frac{p(D, \theta)}{p(D)} \quad (40)$$

$$= \begin{cases} \frac{(N+K)b^N}{K} p(D, \theta) & (m \leq b) \\ \frac{(N+K)m^{N+K}}{Kb^K} p(D, \theta) & (m > b) \end{cases} \quad (41)$$

$$= \begin{cases} \frac{(N+K)b^{N+K}}{\theta^{N+K+1}} I(\theta \geq \max(D, b)) & (m \leq b) \\ \frac{(N+K)m^{N+K}}{\theta^{N+K+1}} I(\theta \geq \max(D, b)) & (m > b) \end{cases} \quad (42)$$

$$= (N+K) \{\max(D, b)\}^{N+K} \theta^{-(N+K+1)} I(\theta \geq \max(D, b)) \quad (43)$$

$$= \text{Pareto}(\theta|N+K, \max(D, b)) \quad (44)$$

Exercise 3.10 Taxicab (tramcar) problem

(a)

$$p(\theta|\{100\}) = \text{Pareto}(\theta|1, 100) \quad (45)$$

$$= 100\theta^{-2} I(\theta \geq 100) \quad (46)$$

(b)

mean

not exist

mode

100

median

$$\int_{100}^x 100\theta^{-2}I(\theta \geq 100)d\theta = \frac{1}{2} \quad (47)$$

$$100 \int_{100}^x \theta^{-2}d\theta = \frac{1}{2} \quad (48)$$

$$-100[\frac{1}{\theta}]_{100}^x = \frac{1}{2} \quad (49)$$

$$\frac{1}{x} - \frac{1}{100} = -\frac{1}{200} \quad (50)$$

$$x = 200 \quad (51)$$

(c)

$$p(D'|D, \alpha) = \int_{\theta} p(D'|\theta)p(\theta|D, \alpha)d\theta \quad (52)$$

$$= \int_{\theta} U(x|0, \theta)\text{Pareto}(\theta|1, m)d\theta \quad (53)$$

$$= \int_{\theta} \frac{1}{\theta} I[0 \leq x \leq \theta] m\theta^{-2} I(\theta \geq m) d\theta \quad (54)$$

$$= m \int_{\theta} \theta^{-3} I[0 \leq x \leq \theta] I[\theta \geq m] d\theta \quad (55)$$

$$= m[-\frac{1}{2}\theta^{-2}]_{\max(x, m)}^{\infty} \quad (56)$$

$$= \frac{m}{2\{\max(x, m)\}^2} \quad (57)$$

(d)

$$p(x = 100|D, \alpha) = \frac{1}{200} \quad (58)$$

$$p(x = 50|D, \alpha) = \frac{1}{200} \quad (59)$$

$$p(x = 150|D, \alpha) = \frac{100}{2 \cdot 150^2} = \frac{1}{450} \quad (60)$$

(e)

- More observations, more accurate.
- Thy hyper-parameter of prior should be set by other information.

Exercise 3.11 Bayesian analysis of the exponential distribution

(a)

$$\log p(D|\theta) = N \log \theta - \theta \sum_{i=1}^N x_i \quad (61)$$

$$\frac{d}{d\theta} \log p(D|\theta) = \frac{N}{\theta} - \sum_{i=1}^N x_i \quad (62)$$

$$\hat{\theta}_{\text{MLE}} = \frac{1}{\bar{x}} \quad (63)$$

(b)

$$\hat{\theta}_{\text{MLE}} = \frac{1}{\frac{5+6+4}{3}} \quad (64)$$

$$= 1/5 \quad (65)$$

(c)

$$E[\theta] = \int_0^{\infty} \theta p(\theta) d\theta \quad (66)$$

$$= \lambda \theta e^{-\lambda \theta} d\theta \quad (67)$$

$$= \frac{1}{\lambda} \quad (68)$$

$$\hat{\lambda} = 3 \quad (69)$$

If we use hint,

$$p(\theta) \propto \text{Ga}(\theta|1, \lambda) \quad (70)$$

$$E[\theta] = \frac{1}{\lambda} \quad (71)$$

(d)

$$p(\theta|D, \hat{\lambda}) \propto p(D|\theta)p(\theta|\hat{\lambda}) \quad (72)$$

$$= \theta^N \exp(-\theta \sum_{i=1}^N x_i) \hat{\lambda} \exp(-\hat{\lambda}\theta) \quad (73)$$

$$= \hat{\lambda} \theta^N \exp(-\theta(\hat{\lambda} + \sum_{i=1}^N x_i)) \quad (74)$$

$$\propto \text{Ga}(\theta|N+1, \hat{\lambda} + \sum_{i=1}^N x_i) \quad (75)$$

(e)

The exponential prior is NOT conjugate to the exponential likelihood.

(The Gamma prior is conjugate to the exponential likelihood.)

(f)

$$E[\theta|D, \hat{\lambda}] = \frac{N+1}{\hat{\lambda} + \sum_{i=1}^N x_i} \quad (76)$$

(g)

$\hat{\lambda}$ can be thought as pseudo sample of machine time. If the expert is collect, this example is reasonable.