Homework 3

Hoang Chu

February 13, 2024

Problem 1.

Solution:

1. Finding the mean of θ:

   Since θ follows a continuous distribution, from the definition of $E[\theta]$:

$$E[\theta] = \int_0^1 \theta \left(\mathbb{P}(\theta|a,b)\right) d\theta$$

$$\leftrightarrow E[\theta] = \int_0^1 \theta(\frac{1}{B(a,b)}\theta^{a-1}(1-\theta)^{b-1}))d\theta$$

$$\leftrightarrow E[\theta] = \frac{1}{B(a,b)}\theta^{a-1}(1-\theta)^{b-1}\int_0^1 \theta(1-\theta)^{b-1}d\theta$$

$$\leftrightarrow E[\theta] = \frac{B(a+1,b)}{B(a,b)}$$

$$\leftrightarrow E[\theta] = \frac{a\Gamma(a)\Gamma(b)}{(a+b)\Gamma(a+b)} \cdot \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} = \boxed{\frac{a}{a+b}}$$

2. Finding the mode of θ: The mode of θ is the value having the highest probability in its distribution, meaning the derivative of its probability distribution w.r.t $\theta = 0$. Hence, we have:

$$\frac{d(\mathbb{P}(\theta,a,b)}{d\theta} = 0$$

$$\leftrightarrow (a-1)\theta^{a-2}(1-\theta)^{b-1} - (b-1)\theta^{a-1}(1-\theta)^{b-2} = 0$$

$$\leftrightarrow (a-1)\theta^{a-2}(1-\theta)^{b-1} = (b-1)\theta^{a-1}(1-\theta)^{b-2}$$

   Since Beta distribution is defined within the $(0,1)$ interval, $\theta \neq 0$ and $1-\theta \neq 0$, meaning we can divide both sides by $((\theta)(1-\theta))^{a-2}$

$$\rightarrow (a-1)(1-\theta) = (b-1)\theta$$

$$\leftrightarrow (a+b-2)\theta = a-1$$

$$\leftrightarrow \theta = \boxed{\frac{a-1}{a+b-2}}$$

3. Finding the variance of θ: Since $Var(\theta) = E[\theta^2] - E[\theta]^2$, and we knew $E[\theta]$ from (1), we can find $E[\theta^2]$ in a similar manner:

$$E[\theta^2] = \int_0^1 \theta^2 \mathbb{P}(\theta, a, b) d\theta$$

$$\leftrightarrow E[\theta^2] = \frac{1}{B(a,b)} \int_0^1 \theta a + 1(1-\theta)^{b-1} d\theta$$

$$\leftrightarrow E[\theta^2] = \frac{B(a+2,b)}{B(a,b)}$$

$$\leftrightarrow E[\theta^2] = \frac{a(a+1)\Gamma(a)\Gamma(b)}{(a+b)(a+b+1)\Gamma(a+b)} \cdot \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} = \frac{a(a+1)}{(a+b)(a+b+1)}$$

Having $E[\theta^2], E[\theta]$, variance of θ is:

$$Var(\theta) = \frac{a(a+1)}{(a+b)(a+b+1)} - \frac{a^2}{(a+b)^2}$$

$$\leftrightarrow Var(\theta) = \frac{a^3 + a^2 b + a^2 + ab - (a^3 + a^2 b + a^2)}{(a+b+1)(a+b)^2} = \boxed{\frac{ab}{(a+b+1)(a+b)^2}}$$

Problem 2.

Solution: The exponential family is in the form:

$$\mathbb{P}((\mathbf{y}, n) = a((\mathbf{y})e^{n^T T((\mathbf{y}) - a(n)}$$

To show that the multinomial distribution is in the exponential family, we can try to fit as much as we can:

$$Cat(\mathbf{x}|\mu) = e^{log(\prod_{i=1}^K \mu_i^{x_i})}$$

$$\leftrightarrow Cat(\mathbf{x}|\mu) = e^{\sum_{i=1}^K log(\mu_i^{x_i})} = ^{\sum_{i=1}^K x_i log(\mu_i)}$$

$$\leftrightarrow Cat(\mathbf{x}|\mu) = e^{\sum_{i=1}^{K-1} x_i log(\frac{\mu_i}{\mu_K}) + log(\mu_k)}$$

Therefore, we ca have the vector n be $[log(\frac{\mu_1}{\mu_k)} \dots log(\frac{\mu_{k-1}}{\mu_k})]^T$

Since $\mu_i = \mu_k \cdot e^n$, we have:

$$\mu_k = 1 - \sum_{i=1}^{K-1} \mu_i = 1 - \sum_{i=1}^{K-1} \mu_k e^n$$

Hence, we can assign $b(n) = 1$, $T(\mathbf{x}) = \mathbf{x}$, and $a(n) = -log(\mu_k) = log(1 + \sum_{i=1}^{K-1} e^n)$. Since every component of the exponential family matches with the distribution, we can conclude that $Cat(\mathbf{x}|\mu)$ belongs to the exponential family. If we assign μ to the Softmax function, we will have the Softmax regression, which implies the generalized linear model of this distribution is the same as Softmax regression. □