Math189R SP19
Homework 3
Monday, Feb 18, 2019

Feel free to work with other students, but make sure you write up the homework and code on your own (no copying homework *or* code; no pair programming). Feel free to ask students or instructors for help debugging code or whatever else, though.

1 (**Murphy 2.16**) Suppose $\theta \sim \text{Beta}(a, b)$ such that

$$\mathbb{P}(\theta; a, b) = \frac{1}{B(a,b)}\theta^{a-1}(1-\theta)^{b-1} = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}$$

where $B(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$ is the Beta function and $\Gamma(x)$ is the Gamma function. Derive the mean, mode, and variance of θ.

To derive the mean, undertake these steps: Determine the expected value via its integral representation. Find the substitution with the Beta distribution's density function. Derive the formula by extracting the constant $\frac{1}{B(a,b)}$. Recognize the integral as a alternate representation of the Beta function. Convert the Beta function into Gamma function notation. Use Gamma function expression by utilizing the recursive property $\Gamma(n+1) = n\Gamma(n)$ and subsequently,
Eliminate corresponding terms:

$$B(a,b) = \int_0^1 \theta^{a-1}(1-\theta)^{b-1} d\theta = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

Thus, the calculation transforms as follows:

$$\begin{aligned}
\mathbb{E}[\theta] &= \int_0^1 \theta \mathbb{P}(\theta; a, b)\, d\theta \\
&= \int_0^1 \theta \left(\frac{1}{B(a,b)}\theta^{a-1}(1-\theta)^{b-1} \right) d\theta \\
&= \frac{1}{B(a,b)} \int_0^1 \theta^{a}(1-\theta)^{b-1}\, d\theta \\
&= \frac{B(a+1,b)}{B(a,b)} \\
&= \frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)} \cdot \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \\
&= \frac{a\Gamma(a)\Gamma(b)}{(a+b)\Gamma(a+b)} \cdot \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \\
&= \frac{a}{a+b}
\end{aligned}$$

In pursuit of the variance:

$$\mathrm{Var}[\theta] = \mathbb{E}[(\theta - \mathbb{E}[\theta])^2] = \mathbb{E}[\theta^2] - (\mathbb{E}[\theta])^2 \tag{1}$$

Employing a similar technique as before, we compute $\mathbb{E}[\theta^2]$:

$$\begin{aligned}
\mathbb{E}[\theta^2] &= \int_0^1 \theta^2 \left(\frac{1}{B(a,b)} \theta^{a-1} (1-\theta)^{b-1} \right) d\theta \\
&= \frac{1}{B(a,b)} \int_0^1 \theta^{a+1} (1-\theta)^{b-1} d\theta \\
&= \frac{B(a+2,b)}{B(a,b)} \\
&= \frac{\Gamma(a+2)\Gamma(b)}{\Gamma(a+b+2)} \cdot \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \\
&= \frac{a(a+1)\Gamma(a)\Gamma(b)}{(a+b)(a+b+1)\Gamma(a+b)} \cdot \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \\
&= \frac{a(a+1)}{(a+b)(a+b+1)} \\
&= \frac{a(a+1)(a+b) - a^2(a+b+1)}{(a+b)^2(a+b+1)} \\
&= \frac{a^3 + a^2b + a^2 + ab - a^3 - a^2b - a^2}{(a+b)^2(a+b+1)} \\
&= \frac{ab}{(a+b)^2(a+b+1)}
\end{aligned}$$

Now we proceed to calculate the Mode:

$$\begin{aligned}
&\nabla_\theta p(\theta; a, b) = \nabla_\theta [\theta^{a-1}(1-\theta)^{b-1}] = 0 \\
&= (a-1)\theta^{a-2}(1-\theta)^{b-1} - (b-1)\theta^{a-1}(1-\theta)^{b-2} = 0 \\
&= (a-1)\theta^{a-2}(1-\theta)^{b-1} = (b-1)\theta^{a-1}(1-\theta)^{b-2} \\
&= (a-1)(1-\theta) = (b-1)\theta \\
&= (a+b-2)\theta = a-1 \\
&\theta^* = \frac{a-1}{a+b-2}
\end{aligned}$$

■

2 (**Murphy 9**) Show that the multinoulli distribution

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

is in the exponential family and show that the generalized linear model corresponding to this distribution is the same as multinoulli logistic regression (softmax regression).

It is understood that this distribution is represented as:

$$P(Y;\eta) = b(y) \exp\left(\eta^T T(y) - a(\eta)\right)$$

By using properties of logarithms, our calculation becomes:

$$\log(ab) = \log(a) + \log(b), \quad \log(a^b) = b \log(a)$$

To demonstrate that the multinomial distribution belongs to the exponential family, we apply the above property:

$$\text{Cat}(x|\mu) = \prod_{i=1}^{K} \mu_i^{x_i} = \exp\left[\log\left(\prod_{i=1}^{K} \mu_i^{x_i}\right)\right]$$
$$= \exp\left(\sum_{i=1}^{K} \log(\mu_i^{x_i})\right)$$
$$= \exp\left(\sum_{i=1}^{K} x_i \log(\mu_i)\right)$$

Given that $\sum_{i=1}^{K} \mu_i = 1$ and $\sum_{i=1}^{K} x_i = 1$, we define the first $K-1$ terms, as the last terms x_K and μ_K will be inherently determined:

$$\mu_K = 1 - \sum_{i=1}^{K-1} \mu_i, \quad x_K = 1 - \sum_{i=1}^{K-1} x_i$$

This allows us to divide our summation as follows:

$$\text{Cat}(x|\mu) = \exp\left(\sum_{k=1}^{K} x_k \log(\mu_k)\right) = \exp\left(\sum_{i=1}^{K-1} x_i \log(\mu_i) + x_K \log(\mu_K)\right)$$
$$= \exp\left(\sum_{i=1}^{K-1} x_i \log(\mu_i) + \left(1 - \sum_{i=1}^{K-1} x_i\right) \log(\mu_K)\right)$$
$$= \exp\left(\sum_{i=1}^{K-1} x_i \log(\mu_i) - \log(\mu_K) \sum_{i=1}^{K-1} x_i\right) + \log(\mu_K)$$
$$= \exp\left(\sum_{i=1}^{K-1} x_i \log\left(\frac{\mu_i}{\mu_K}\right) + \log(\mu_K)\right)$$

Introducing the vector η as

$$\eta = \begin{bmatrix} \log\left(\frac{\mu_1}{\mu_K}\right) \\ \vdots \\ \log\left(\frac{\mu_{K-1}}{\mu_K}\right) \end{bmatrix}$$

enables us to express $\mu_i = \mu_K e^{\eta_i}$, leading to the substitution:

$$\mu_K = 1 - \sum_{i=1}^{K-1} \mu_i = 1 - \mu_K \sum_{i=1}^{K-1} e^{\eta_i}$$

$$= \frac{1}{1 + \sum_{i=1}^{K-1} e^{\eta_i}}$$

$$\therefore \mu_i = \mu_K e^{\eta_i} = \frac{e^{\eta_i}}{1 + \sum_{i=1}^{K-1} e^{\eta_i}}$$

Writing the distribution in the form of exponential family as $\mathrm{Cat}(x|\mu) = \exp(\eta^T x - a(\eta))$:

$$b(\eta) = 1$$

$$T(x) = x$$

$$a(\eta) = -\log(\mu_K) = \log\left(1 + \sum_{i=1}^{K-1} e^{\eta_i}\right)$$

$\therefore$ The distribution $\mathrm{Cat}(x|\mu)$ is in the exponential family. And $\mu = S(\eta)$, where $S(\eta)$ is the softmax function. ∎